Joint k-step analysis of Orthogonal Matching Pursuit and Orthogonal Least Squares

Charles Soussen, Rémi Gribonval, Jérôme Idier, Cédric Herzet

Joint k-step analysis of Orthogonal Matching Pursuit and Orthogonal Least Squares

Charles Soussen*, Rémi Gribonval, Jérôme Idier, and Cédric Herzet

***Abstract*—Tropp's analysis of Orthogonal Matching Pursuit (OMP) using the Exact Recovery Condition (ERC) [1] is extended to a first exact recovery analysis of Orthogonal Least Squares (OLS). We show that when the ERC is met, OLS is guaranteed to exactly recover the unknown support in at most k iterations. Moreover, we provide a closer look at the analysis of both OMP and OLS when the ERC is not fulfilled. The existence of dictionaries for which some subsets are never recovered by OMP is proved. This phenomenon also appears with basis pursuit where support recovery depends on the sign patterns, but it does not occur for OLS. Finally, numerical experiments show that none of the considered algorithms is uniformly better than the other but for correlated dictionaries, guaranteed exact recovery may be obtained after fewer iterations for OLS than for OMP.**

***Index Terms*—ERC exact recovery condition; Orthogonal Matching Pursuit; Orthogonal Least Squares; Order Recursive Matching Pursuit; Optimized Orthogonal Matching Pursuit; forward selection.**

I. Introduction

CLASSICAL greedy subset selection algorithms include, by increasing order of complexity: Matching Pursuit (MP) [2], Orthogonal Matching Pursuit (OMP) [3] and Orthogonal Least Squares (OLS) [4, 5]. OLS is indeed relatively expensive in comparison with OMP since OMP performs one linear inversion per iteration whereas OLS performs as many linear inversions as there are non-active atoms. We refer the reader to the technical report [6] for a comprehensive review on the difference between OMP and OLS.

OLS is referred to using many other names in the literature. It is known as forward selection in statistical regression [7] and as the greedy algorithm [5], Order Recursive Matching Pursuit (ORMP) [8] and Optimized Orthogonal Matching Pursuit (OOMP) [9] in the signal processing literature, all these algorithms being actually the same. It is worth noticing that the above-mentioned algorithms were introduced by following either an optimization [4, 7] or an orthogonal projection methodology [5], or both [8, 9]. In the optimization viewpoint, the atom yielding the largest decrease of the approximation error is selected. This leads to a greedy sub-optimal algorithm dedicated to the minimization of the approximation error. In the orthogonal projection viewpoint, the atom selection rule is defined as an extension of the OMP rule: the data vector and the dictionary atoms are being projected onto the subspace that is orthogonal to the span of the active atoms, and the *normalized* projected atom having the largest inner product with the data residual is selected. As the number of active atoms increases by one at any iteration, the projections are done on a subspace whose dimension is decreasing.

A. Main objective of the paper

Our primary goal is to address the OLS exact recovery analysis from noise-free data and to investigate the connection between the OMP and OLS exact recovery conditions. In the literature, much attention was paid to the exact recovery analysis of sparse algorithms that are faster than OLS, *e.g.*, thresholding algorithms and simpler greedy algorithms like OMP [10]. But to the best of our knowledge, no exact recovery result is available for OLS. In their recent paper [11], Davies and Eldar mention this issue and state that the relation between OMP and OLS remains unclear.

B. Existing results for OMP

Our starting point is the existing k-step analysis of OMP whose structure is somewhat close to OLS. The notion of k-step solution property was defined in [12]:

C. Soussen is with the Centre de Recherche en Automatique de Nancy (CRAN, UMR 7039, Université de Lorraine, CNRS). Campus Sciences, B.P. 70239, F-54506 Vandœuvre-lès-Nancy, France (e-mail: Charles.Soussen@cran.uhp-nancy.fr.) This work was carried out in part while C. Soussen was visiting IRCCyN during the academic year 2010-2011 with the financial support of CNRS.

R. Gribonval and C. Herzet are with INRIA Rennes - Bretagne Atlantique, Campus de Beaulieu, F-35042 Rennes Cedex, France (e-mail: Remi.Gribonval@inria.fr; Cedric.Herzet@inria.fr). R. Gribonval acknowledges the partial support of the European Union's FP7-FET program, SMALL project, under grant agreement n° 225913.

J. Idier is with the Institut de Recherche en Communications et Cybernétique de Nantes (IRCCyN, UMR CNRS 6597), BP 92101, 1 rue de la Noë, 44321 Nantes Cedex 3, France (e-mail: Jerome.Idier@irccyn.ec-nantes.fr).

"any vector with at most k nonzeros can be recovered from the related noise-free observation in at most k iterations." The k-step property will also be referred to as the "exact support recovery" in the following. Exact recovery studies of OMP rely on alternate methodologies.

Tropp's Exact Recovery Condition (ERC) [1] is a necessary and sufficient condition of exact support recovery in a worst case analysis. On the one hand, if a subset of k atoms satisfies the ERC, then it can be recovered from any linear combination of the k atoms in at most k steps. On the other hand, when the ERC is not satisfied, one can generate a counterexample (*i.e.,* a specific combination of the k atoms) for which OMP fails, *i.e.*, OMP selects a wrong atom during its first k iterations. Specifically, the atom selected in the *first* iteration is a wrong one.

Davenport and Wakin [13] used another analysis to show that OMP yields exact support recovery under certain Restricted Isometry Property (RIP) assumptions, and several improvements of their condition were proposed more recently [14, 15]. Actually, the ERC necessarily holds when the latter conditions are fulfilled since the ERC is a sufficient and worst case necessary condition of exact recovery.

C. Generalization of Tropp's condition

We propose to extend Tropp's condition to OLS. We remark that the very first iteration of OLS is identical to that of OMP: the first selected atom is the one whose inner product with the input vector is maximal. Therefore, when the ERC does not hold, the counterexample for which the first iteration of OMP fails also yields a failure of the first iteration of OLS. Hence one cannot expect to derive an exact recovery condition for OLS that would be weaker than the ERC at the first iteration. We show that the ERC indeed ensures the success of OLS.

We further address the case where the ERC does not hold, *i.e.*, the first iteration of OMP/OLS is not guaranteed to always succeed but nevertheless succeeds for a given vector. In practice, even for non random dictionaries, this phenomenon is likely to occur since the ERC is a worst case necessary condition. The purpose of a large part of the paper is specifically to analyze what is going on in the remaining iterations for these vectors. With ℓ_1 minimization, the situation is clearer because support recovery depends on the sign patterns [16, Theorem 2] and one can predict whether a specific vector will be recovered independently of the support amplitudes. For greedy algorithms, things are more tricky and it is one of the purpose of the paper to analyze this. We introduce weaker conditions than the ERC which guarantee that an exact support recovery will occur in the subsequent iterations. These extended recovery conditions coincide with the ERC at the first iteration but differ from it afterwards.

Our main results state that:

- The ERC is a sufficient condition of exact recovery for OLS in at most k steps (Theorem 2).
- When the early iterations of OMP/OLS have all succeeded, we derive two sufficient conditions, named ERC-OMP and ERC-OLS, for the recovery of the remaining true atoms (Theorem 3). This result is a $(k-q)$-step property, where q stands for the number of iterations which have been already performed.
- Moreover, we show that our conditions are, in some sense, necessary (Theorems 4 and 5).

The criteria we provide might not necessarily be directly useful for practitioners working in the field. In fact, just as many other theoretical success guarantees, they are rather "motivational": by proving that the considered algorithms are guaranteed to perform well in a restricted regime, they strengthen our confidence that the heuristics behind the algorithms are reasonably grounded. Practitioners know that the algorithms indeed work much beyond the considered restricted regime, but proving this fact would typically require probabilistic arguments, based on models of random dictionary or random input signals [17, 18]. Despite their potential interest, the theoretical results that can be foreseen in this spirit would be highly dependent on the adequacy of such models to the actual distribution of data from the real world.

D. Organization of the paper

In Section II, we recall the principle of OMP and OLS and their interpretation in terms of orthogonal projections. Then, we properly define the notions of successful support recovery and support recovery failure. Section III is dedicated to the analysis of OMP and OLS at any iteration where the most technical developments and proofs are omitted for readability reasons. These important elements can be found in the appendix section A. In Section IV, we show using Monte Carlo simulations that there is no systematic implication between the ERC-OMP and ERC-OLS conditions but we exhibit some elements of discrimination in favor of OLS.

II. Notations and prerequisites

The following notations will be used in this paper. $\langle\, .\, ,\, .\, \rangle$ refers to the inner product between vectors, and $\|\, .\, \|$ and $\|\, .\, \|_1$ stand for the Euclidean norm and the ℓ_1 norm, respectively. $.^\dagger$ denotes the pseudo-inverse of a matrix. For a full rank and undercomplete matrix, we have $\boldsymbol{X}^\dagger = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t$ where $.^t$ stands for the matrix transposition. When $\boldsymbol{X}$ is overcomplete, $\mathrm{spark}(\boldsymbol{X})$ denotes the minimum number of columns from $\boldsymbol{X}$ that are linearly dependent [19]. The letter $\mathcal{Q}$ denotes some subset of the column indices, and $\boldsymbol{X}_\mathcal{Q}$ is the submatrix of $\boldsymbol{X}$ gathering the columns indexed by $\mathcal{Q}$. Finally, $\boldsymbol{P}_\mathcal{Q} = \boldsymbol{X}_\mathcal{Q}\boldsymbol{X}_\mathcal{Q}^\dagger$ and $\boldsymbol{P}_\mathcal{Q}^\perp = \boldsymbol{I} - \boldsymbol{P}_\mathcal{Q}$ denote the orthogonal projection operators on $\mathrm{span}(\boldsymbol{X}_\mathcal{Q})$ and $\mathrm{span}(\boldsymbol{X}_\mathcal{Q})^\perp$, where $\mathrm{span}(\boldsymbol{X})$ stands for the column span of $\boldsymbol{X}$, $\mathrm{span}(\boldsymbol{X})^\perp$ is the orthogonal complement of $\mathrm{span}(\boldsymbol{X})$ and $\boldsymbol{I}$ is the identity matrix whose dimension is equal to the number of rows in $\boldsymbol{X}$.

A. *Subset selection*

Let $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$ denote the dictionary gathering normalized atoms $\boldsymbol{a}_i \in \mathbb{R}^m$. $\boldsymbol{A}$ is a matrix of size $m \times n$. Assuming that the atoms are normalized is actually not necessary for OLS as the behavior of OLS is unchanged whether the atoms are normalized or not [6]. On the contrary, OMP is highly sensitive to the normalization of atoms since its selection rule involves the inner products between the current residual and the non-selected atoms.

We consider a subset $\mathcal{Q}^\star$ of $\{1, \ldots, n\}$ of cardinality $k \triangleq \mathrm{Card}\,[\mathcal{Q}^\star] < \min(m, n)$ and study the behavior of OMP and OLS *for all* inputs $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$, *i.e.*, for any combination $\boldsymbol{y} = \boldsymbol{A}_{\mathcal{Q}^\star}\boldsymbol{t}$ where the submatrix $\boldsymbol{A}_{\mathcal{Q}^\star}$ is of size $m \times k$ and the weight vector $\boldsymbol{t} \in \mathbb{R}^k$. The k atoms $\{\boldsymbol{a}_i,\, i \in \mathcal{Q}^\star\}$ indexed by $\mathcal{Q}^\star$ will be referred to as the "true" atoms while for the remaining ("wrong") atoms $\{\boldsymbol{a}_j,\, j \notin \mathcal{Q}^\star\}$, we will use the subscript notation j. The forward greedy algorithms considered in this paper start from the empty support and select a new atom per iteration. At intermediate iterations $q \in \{0, \ldots, k-1\}$, we denote by $\mathcal{Q}$ the current support (with $\mathrm{Card}\,[\mathcal{Q}] = q$).

Throughout the paper, we make the general assumption that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. Note that the representation $\boldsymbol{y} = \boldsymbol{A}_{\mathcal{Q}^\star}\boldsymbol{t}$ is not guaranteed to be unique under this assumption: there may be another k-term representation $\boldsymbol{y} = \boldsymbol{A}_{\mathcal{Q}'}\boldsymbol{t}'$ where $\boldsymbol{A}_{\mathcal{Q}'}$ includes some wrong atoms $\boldsymbol{a}_j$. The stronger assumption $\mathrm{spark}(\boldsymbol{A}) > 2k$ is a necessary and sufficient condition for uniqueness of any k-term representation [19]. Therefore, when $\mathrm{spark}(\boldsymbol{A}) > 2k$, the selection of a wrong atom by a greedy algorithm disables a k-term representation of $\boldsymbol{y}$ in k steps [1]. We make the weak assumption that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank because it is sufficient to elaborate our exact recovery conditions under which no wrong atom is selected in the first k iterations.

B. *OMP and OLS algorithms*

The common feature between OMP and OLS is that they both perform an orthogonal projection whenever the support $\mathcal{Q}$ is updated: the data approximation reads $\boldsymbol{P}_\mathcal{Q}\boldsymbol{y}$ and the residual error is defined by

$$\boldsymbol{r}_\mathcal{Q} \triangleq \boldsymbol{y} - \boldsymbol{P}_\mathcal{Q}\boldsymbol{y} = \boldsymbol{P}_\mathcal{Q}^\perp\boldsymbol{y}.$$

Let us now recall how the selection rule of OLS differs from that of OMP.

At each iteration of OLS, the atom $\boldsymbol{a}_\ell$ yielding the minimum least-square error $\|\boldsymbol{r}_{\mathcal{Q}\cup\{\ell\}}\|^2$ is selected:

$$\ell^{\mathrm{OLS}} \in \arg\min_{i \notin \mathcal{Q}} \|\boldsymbol{r}_{\mathcal{Q}\cup\{i\}}\|^2$$

and $n - \mathrm{Card}\,[\mathcal{Q}]$ least-square problems are being solved to compute $\|\boldsymbol{r}_{\mathcal{Q}\cup\{i\}}\|^2$ for all $i \notin \mathcal{Q}$ ([1]) [4]. On the contrary, OMP adopts the simpler rule

$$\ell^{\mathrm{OMP}} \in \arg\max_{i \notin \mathcal{Q}} |\langle \boldsymbol{r}_\mathcal{Q}, \boldsymbol{a}_i\rangle|$$

to select the new atom $\boldsymbol{a}_\ell$ and then solves only one least-square problem to update $\boldsymbol{r}_{\mathcal{Q}\cup\{\ell\}}$ [6]. Depending on the application, the OMP and OLS stopping rules can involve a maximum number of atoms and/or a residual threshold. Note that when the data are noise-free (they read as $\boldsymbol{y} = \boldsymbol{A}_{\mathcal{Q}^\star}\boldsymbol{t}$) and no wrong atom is selected, the squared error $\|\boldsymbol{r}_\mathcal{Q}\|^2$ is equal to 0 after at most k iterations. Therefore, we will consider no more than k iterations in the following.

C. *Geometric interpretation*

A geometric interpretation in terms of orthogonal projections will be useful for deriving recovery conditions. It is essentially inspired by the technical report of Blumensath and Davies [6] and by Davenport and Wakin's analysis of OMP under the RIP assumption [13].

We introduce the notation $\tilde{\boldsymbol{a}}_i = \boldsymbol{P}_\mathcal{Q}^\perp\boldsymbol{a}_i$ for the projected atoms onto $\mathrm{span}(\boldsymbol{A}_\mathcal{Q})^\perp$ where for simplicity, the dependence upon $\mathcal{Q}$ is omitted. When there is a risk of confusion, we will use $\tilde{\boldsymbol{a}}_i^\mathcal{Q}$ instead of $\tilde{\boldsymbol{a}}_i$. Notice that $\tilde{\boldsymbol{a}}_i = \boldsymbol{0}$ *if and only if* $\boldsymbol{a}_i \in \mathrm{span}(\boldsymbol{A}_\mathcal{Q})$. In particular, $\tilde{\boldsymbol{a}}_i = \boldsymbol{0}$ for $i \in \mathcal{Q}$. Finally, we define the normalized vectors

$$\tilde{\boldsymbol{b}}_i = \begin{cases} \tilde{\boldsymbol{a}}_i/\|\tilde{\boldsymbol{a}}_i\| & \text{if } \tilde{\boldsymbol{a}}_i \neq \boldsymbol{0}, \\ \boldsymbol{0} & \text{otherwise.} \end{cases}$$

Again, we will use $\tilde{\boldsymbol{b}}_i^\mathcal{Q}$ when there is a risk of confusion.

We now emphasize that the projected atoms $\tilde{\boldsymbol{a}}_i$ (or $\tilde{\boldsymbol{b}}_i$) play a central role in the analysis of both OMP and OLS.

[1]Our purpose is not to focus on the OLS implementation. However, let us just mention that in the typical implementation, the least-square problems are solved recursively using the Gram Schmidt orthonormalization procedure [4].

Because the residual $\boldsymbol{r}_\mathcal{Q} = \boldsymbol{P}_\mathcal{Q}^\perp \boldsymbol{y}$ lays in $\mathrm{span}(\boldsymbol{A}_\mathcal{Q})^\perp$, $\langle \boldsymbol{r}_\mathcal{Q}, \boldsymbol{a}_i \rangle = \langle \boldsymbol{r}_\mathcal{Q}, \tilde{\boldsymbol{a}}_i \rangle$ and the OMP selection rule rereads:

$$\ell^{\mathrm{OMP}} \in \arg\max_{i \notin \mathcal{Q}} |\langle \boldsymbol{r}_\mathcal{Q}, \tilde{\boldsymbol{a}}_i \rangle| \tag{1}$$

whereas for OLS, minimizing $\|\boldsymbol{r}_{\mathcal{Q}\cup\{i\}}\|^2$ with respect to $i \notin \mathcal{Q}$ is equivalent to maximizing $\|\boldsymbol{r}_\mathcal{Q}\|^2 - \|\boldsymbol{r}_{\mathcal{Q}\cup\{i\}}\|^2 = \langle \boldsymbol{r}_\mathcal{Q}, \tilde{\boldsymbol{b}}_i \rangle^2$ (see *e.g.,* [9] for a complete calculation):

$$\ell^{\mathrm{OLS}} \in \arg\max_{i \notin \mathcal{Q}} |\langle \boldsymbol{r}_\mathcal{Q}, \tilde{\boldsymbol{b}}_i \rangle|. \tag{2}$$

We notice that (1) and (2) only rely on the vectors $\boldsymbol{r}_\mathcal{Q}$ and $\tilde{\boldsymbol{a}}_i$ belonging to the subspace $\mathrm{span}(\boldsymbol{A}_\mathcal{Q})^\perp$. OMP maximizes the inner product $|\langle \boldsymbol{r}_\mathcal{Q}, \tilde{\boldsymbol{a}}_i \rangle|$ whereas OLS minimizes the angle between $\boldsymbol{r}_\mathcal{Q}$ and $\tilde{\boldsymbol{a}}_i$ (this difference was already stressed and graphically illustrated in [6]). When the dictionary is close to orthogonal, *e.g.,* for dictionaries satisfying the RIP assumption, this does not make a strong difference since $\|\tilde{\boldsymbol{a}}_i\|$ is close to 1 for all atoms [13]. But in the general case, $\|\tilde{\boldsymbol{a}}_i\|$ may have wider variations between 0 and 1 leading to substantial differences between the behavior of OMP and OLS.

D. Definition of successful recovery and failure

Throughout the paper, we will use the common acronym Oxx in statements that apply to both OMP and OLS. Moreover, we define the unifying notation:

$$\tilde{\boldsymbol{c}}_i \triangleq \begin{cases} \tilde{\boldsymbol{a}}_i & \text{for OMP,} \\ \tilde{\boldsymbol{b}}_i & \text{for OLS.} \end{cases}$$

We first stress that in special cases where the Oxx selection rule yields multiple solutions including a wrong atom, *i.e.,* when

$$\max_{i \in \mathcal{Q}^\star \setminus \mathcal{Q}} |\langle \boldsymbol{r}_\mathcal{Q}, \tilde{\boldsymbol{c}}_i \rangle| = \max_{j \notin \mathcal{Q}^\star} |\langle \boldsymbol{r}_\mathcal{Q}, \tilde{\boldsymbol{c}}_j \rangle|, \tag{3}$$

we consider that Oxx automatically makes the wrong decision. Tropp used this convention for OMP and showed that when the upper bound on his ERC condition (see Section III-A) is reached, the limit situation (3) occurs, hence a wrong atom is selected at the first iteration [1]. Let us now properly define the k-step property for successful support recovery.

Definition 1 *Oxx with $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ as input succeeds if and only if no wrong atom is selected and the residual $\boldsymbol{r}_\mathcal{Q}$ is equal to $\boldsymbol{0}$ after at most k iterations.*

When a successful recovery occurs, the subset $\mathcal{Q}$ yielded by Oxx satisfies $\mathcal{Q}_{\boldsymbol{y}} \subseteq \mathcal{Q} \subseteq \mathcal{Q}^\star$ where $\mathcal{Q}_{\boldsymbol{y}}$ is the subset indexed by the nonzero weights t_i's in the decomposition $\boldsymbol{y} = \boldsymbol{A}_{\mathcal{Q}^\star}\boldsymbol{t}$. When all t_i's are nonzero, $\mathcal{Q}_{\boldsymbol{y}}$ identifies with $\mathcal{Q}^\star$ and a successful recovery cannot occur in less than k iterations.

The word "failure" refers to the exact contrary of successful recovery.

Definition 2 *Oxx with $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ as input fails when at least one wrong atom is selected during the first k iterations. In particular, Oxx fails when* (3) *occurs with* $\boldsymbol{r}_\mathcal{Q} \neq \boldsymbol{0}$.

The notion of successful recovery may be defined in a weaker sense: Plumbley [16, Corollary 4] pointed out that there exist problems for which the ERC fails but nevertheless, a "delayed recovery" occurs after more than k steps, in that a larger support including $\mathcal{Q}^\star$ is found, but all atoms which do not belong to $\mathcal{Q}^\star$ are weighted by 0 in the solution vector. Recently, a delayed recovery analysis of OMP using RIP assumptions was proposed in [20], and then extended to the weak OMP algorithm [21]. In the present paper, no more than k steps are performed, thus delayed recovery is considered as a recovery failure.

III. Overview of our recovery analysis of OMP and OLS

In this section, we present our main concepts and results regarding the sparse recovery guarantees with OLS, their connection with the existing OMP results and the new results regarding OMP. For clarity reasons, we place the technical analysis including most of the proofs in the main appendix section A. Let us first recall Tropp's ERC condition for OMP which is our starting point.

A. Tropp's ERC condition for OMP

Theorem 1 ***[ERC is a sufficient recovery condition for OMP and a necessary condition at the first iteration [1, Theorems 3.1 and 3.10]]*** *If $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank and*

$$\max_{j \notin \mathcal{Q}^\star} \{F_{\mathcal{Q}^\star}(\boldsymbol{a}_j) \triangleq \left\|\boldsymbol{A}_{\mathcal{Q}^\star}^\dagger \boldsymbol{a}_j\right\|_1\} < 1, \qquad \mathrm{ERC}(\boldsymbol{A}, \mathcal{Q}^\star)$$

then OMP succeeds for any input $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$. Furthermore, when ERC($\boldsymbol{A}, \mathcal{Q}^\star$) does not hold, there exists $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ for which some wrong atom is selected at the first iteration of OMP. When $\mathrm{spark}(\boldsymbol{A}) > 2k$, this implies that OMP cannot recover the (unique) k-term representation of $\boldsymbol{y}$.

Note that ERC($\boldsymbol{A}, \mathcal{Q}^\star$) involves the dictionary atoms but not their weights as it results from a worst case analysis: if ERC($\boldsymbol{A}, \mathcal{Q}^\star$) holds, then a successful recovery occurs with $\boldsymbol{y} = \boldsymbol{A}_{\mathcal{Q}^\star}\boldsymbol{t}$ *whatever* $\boldsymbol{t} \in \mathbb{R}^k$.

B. Main theorem

A theorem similar to Theorem 1 applies to OLS.

Theorem 2 ***[ERC is a sufficient recovery condition for OLS and a necessary condition at the first iteration]*** *If $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank and ERC($\boldsymbol{A}, \mathcal{Q}^\star$) holds, then OLS succeeds for any input $\boldsymbol{y} \in \operatorname{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$. Furthermore, when ERC($\boldsymbol{A}, \mathcal{Q}^\star$) does not hold, there exists $\boldsymbol{y} \in \operatorname{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ for which some wrong atom is selected at the first iteration of OLS. When $\operatorname{spark}(\boldsymbol{A}) > 2k$, this implies that OLS cannot recover the (unique) k-term representation of $\boldsymbol{y}$.*

The necessary condition result is obvious since the very first iteration of OLS coincides with that of OMP and the ERC is a worst case necessary condition for OMP. The core of our contribution is the sufficient condition result for OLS. We now introduce the main concepts on which our analysis relies. They also lead to a more precise analysis of OMP from the second iteration.

C. Main concepts

Let us keep in mind that the ERC is a worst case necessary condition *at the first iteration*. But what happens when the ERC is not met but nevertheless, the first q iterations of Oxx select q true atoms ($q < k$)? Can we characterize the exact recovery conditions at the $(q+1)$-th iteration? We will answer to these questions and provide:

1) an extension of the ERC condition to the q-th iteration of OMP;
2) a new necessary and sufficient condition dedicated to the q-th iteration of OLS.

This will allow us to prove Theorem 2 as a special case of the latter condition when $q = 0$.

In the following two paragraphs, we introduce useful notations for a single wrong atom $\boldsymbol{a}_j$ and then define our new exact recovery conditions by considering all wrong atoms together. $\mathcal{Q}$ plays the role of the subset found by Oxx after the first q iterations.

1) Notations related to a single wrong atom: For $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ and $j \notin \mathcal{Q}^\star$, we define:

$$F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) \triangleq \sum_{i \in \mathcal{Q}^\star \setminus \mathcal{Q}} \left| \left(\boldsymbol{A}^\dagger_{\mathcal{Q}^\star} \boldsymbol{a}_j \right)(i) \right| \tag{4}$$

$$F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) \triangleq \sum_{i \in \mathcal{Q}^\star \setminus \mathcal{Q}} \frac{\|\tilde{\boldsymbol{a}}_i\|}{\|\tilde{\boldsymbol{a}}_j\|} \left| \left(\boldsymbol{A}^\dagger_{\mathcal{Q}^\star} \boldsymbol{a}_j \right)(i) \right| \tag{5}$$

when $\tilde{\boldsymbol{a}}_j \neq \boldsymbol{0}$ and $F^{\mathrm{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) = 0$ when $\tilde{\boldsymbol{a}}_j = \boldsymbol{0}$ (we recall that $\tilde{\boldsymbol{a}}_i = \boldsymbol{P}^\perp_{\mathcal{Q}} \boldsymbol{a}_i$ and $\tilde{\boldsymbol{a}}_j = \boldsymbol{P}^\perp_{\mathcal{Q}} \boldsymbol{a}_j$ depend on $\mathcal{Q}$). Up to some manipulations on orthogonal projections, (4) and (5) can be rewritten as follows.

Lemma 1 *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. For $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ and $j \notin \mathcal{Q}^\star$, $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$ and $F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$ also read*

$$F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) = \|\tilde{\boldsymbol{A}}^\dagger_{\mathcal{Q}^\star \setminus \mathcal{Q}} \tilde{\boldsymbol{a}}_j\|_1 \tag{6}$$

$$F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) = \|\tilde{\boldsymbol{B}}^\dagger_{\mathcal{Q}^\star \setminus \mathcal{Q}} \tilde{\boldsymbol{b}}_j\|_1 \tag{7}$$

where the matrices $\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star \setminus \mathcal{Q}} = \{\tilde{\boldsymbol{a}}_i,\, i \in \mathcal{Q}^\star \setminus \mathcal{Q}\}$ and $\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star \setminus \mathcal{Q}} = \{\tilde{\boldsymbol{b}}_i,\, i \in \mathcal{Q}^\star \setminus \mathcal{Q}\}$ of size $m \times (k-q)$ are full rank.

Lemma 1 is proved in Appendix B.

2) ERC-Oxx conditions for the whole dictionary: We define four binary conditions by considering all the wrong atoms together:

$$\max_{j \notin \mathcal{Q}^\star} F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) < 1 \qquad \text{ERC-OMP}(\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q})$$

$$\max_{j \notin \mathcal{Q}^\star} F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) < 1 \qquad \text{ERC-OLS}(\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q})$$

$$\max_{\substack{\mathcal{Q} \subsetneq \mathcal{Q}^\star \\ \operatorname{Card}[\mathcal{Q}] = q}} \max_{j \notin \mathcal{Q}^\star} F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) < 1 \qquad \text{ERC-OMP}(\boldsymbol{A}, \mathcal{Q}^\star, q)$$

$$\max_{\substack{\mathcal{Q} \subsetneq \mathcal{Q}^\star \\ \operatorname{Card}[\mathcal{Q}] = q}} \max_{j \notin \mathcal{Q}^\star} F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) < 1 \qquad \text{ERC-OLS}(\boldsymbol{A}, \mathcal{Q}^\star, q)$$

We will use the common notations $F^{\mathrm{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$, ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) and ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, q$) for statements that are common to both OMP and OLS.

Remark 1 *$F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\emptyset}(\boldsymbol{a}_j)$ and $F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\emptyset}(\boldsymbol{a}_j)$ both reread $F_{\mathcal{Q}^\star}(\boldsymbol{a}_j) = \left\|\boldsymbol{A}^\dagger_{\mathcal{Q}^\star} \boldsymbol{a}_j\right\|_1$ since $\tilde{\boldsymbol{a}}^\emptyset_i$ reduces to $\boldsymbol{a}_i$ which is of unit norm. Thus, ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, \emptyset$) and ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, 0$) all identify with ERC($\boldsymbol{A}, \mathcal{Q}^\star$).*

D. Sufficient conditions of exact recovery at any iteration

The sufficient conditions of Theorems 1 and 2 reread as special cases of the following theorem where $\mathcal{Q} = \emptyset$.

Theorem 3 ***[Sufficient recovery condition for Oxx after q successful iterations]*** *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. If Oxx with $\boldsymbol{y} \in \operatorname{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ as input selects $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ and ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) holds, then Oxx succeeds in at most k steps.*

The following corollary is a straightforward adaptation of Theorem 3 to ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, q$).

Corollary 1 *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. If Oxx with $\boldsymbol{y} \in \operatorname{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ as input selects true atoms during the first $q < k$ iterations and ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, q$) holds, then Oxx succeeds in at most k iterations.*

The key element which enables us to establish Theorem 3 is a recursive relation linking $F^{\mathrm{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$ with

$F^{\rm Oxx}_{\mathcal{Q}^\star,\mathcal{Q}'}(\boldsymbol{a}_j)$ when $\mathcal{Q}$ is increased by one element of $\mathcal{Q}^\star\backslash\mathcal{Q}$, resulting in subset $\mathcal{Q}'$. This leads to the main technical novelty of the paper, stated in Lemma 7 (see Appendix A-A). From the thorough analysis of this recursive relation, we elaborate the following lemma which guarantees the monotonic decrease of $F^{\rm Oxx}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$ when $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ is growing.

Lemma 2 *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. Let $\mathcal{Q} \subsetneq \mathcal{Q}' \subsetneq \mathcal{Q}^\star$. For any $j \notin \mathcal{Q}^\star$,*

$$F^{\rm OMP}_{\mathcal{Q}^\star,\mathcal{Q}'}(\boldsymbol{a}_j) \leqslant F^{\rm OMP}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) \quad (8)$$

$$F^{\rm OLS}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) < 1 \Rightarrow F^{\rm OLS}_{\mathcal{Q}^\star,\mathcal{Q}'}(\boldsymbol{a}_j) \leqslant F^{\rm OLS}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) \quad (9)$$

We refer the reader to Appendix A-A for the proofs of Lemmas 7 and 2, and then Theorem 3.

E. Necessary conditions of exact recovery at any iteration

We recall that the ERC is a worst case necessary condition guaranteed for the selection of a true atom by OMP and OLS in their very first iteration. We provide extended results stating that ERC-Oxx are worst case necessary conditions when the first iterations of Oxx have succeeded, up to a "reachability assumption" defined hereafter, for OMP.

Definition 3 ***[Reachability]*** *$\mathcal{Q}$ is reachable if and only if there exists an input $\boldsymbol{y} = \boldsymbol{A}_\mathcal{Q}\boldsymbol{t}$ where $t_i \neq 0$ for all i, for which Oxx recovers $\mathcal{Q}$ in $\mathrm{Card}\,[\mathcal{Q}]$ iterations. Specifically, the selection rule* (1)-(2) *always yields a unique maximum.*

We start with the OLS condition which is simpler.

1) OLS necessary condition:

Theorem 4 ***[Necessary condition for OLS after q iterations]*** *Let $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ be a subset of cardinality q. Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank and $\mathrm{spark}(\boldsymbol{A}) \geqslant (q+2)$. If ERC-OLS($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) does not hold, then there exists $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ for which OLS selects $\mathcal{Q}$ in the first q iterations and then a wrong atom $j \notin \mathcal{Q}^\star$ at iteration $(q+1)$.*

Theorem 4 is proved in Appendix A-B. An obvious corollary can be obtained by replacing $\mathcal{Q}$ with q akin to the derivation of Corollary 1 from Theorem 3. From now on, such obvious corollaries will not be explicitly stated.

2) Reachability issues: The reader may have noticed that Theorem 4 implies that $\mathcal{Q}$ can be reached by OLS at least for some input $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$. In Appendix A-B, we establish a stronger result:

Lemma 3 (Reachability by OLS) *Any subset $\mathcal{Q}$ with $\mathrm{Card}\,[\mathcal{Q}] \leqslant \mathrm{spark}(\boldsymbol{A}) - 2$ can be reached by OLS with some input $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_\mathcal{Q})$.*

Perhaps surprisingly, this result does not remain valid for OMP although it holds under certain RIP assumptions [13, Theorem 4.1]. We refer the reader to subsection IV-C for a simple counterexample where $\mathcal{Q}$ cannot be reached by OMP not only for $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_\mathcal{Q})$ but also for any $\boldsymbol{y} \in \mathbb{R}^m$.

The same somewhat surprising phenomenon of non-reachability may also occur with ℓ_1 minimization, associated to certain k-faces of the ℓ_1 ball in $\mathbb{R}^n$ whose projection through $\boldsymbol{A}$ yields interior faces [22]. Specifically, for a given $\boldsymbol{x}$ supported by $\mathcal{Q}$, Fuchs' necessary and sufficient condition for exact support recovery from $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ [16, 23] involves the signs of the nonzero amplitudes (denoted by $\boldsymbol{\varepsilon} \triangleq \mathrm{sgn}(\boldsymbol{x}) \in \{-1,1\}^q$) but not their values. Either Fuchs' condition is met *for any* vector having support $\mathcal{Q}$ and signs $\boldsymbol{\varepsilon}$, thus all these vectors will be correctly recovered, or no vector $\boldsymbol{x}$ having support $\mathcal{Q}$ and signs $\boldsymbol{\varepsilon}$ can be recovered. It follows that $\mathcal{Q}$ is non-reachable with ℓ_1 minimization when Fuchs' condition is simultaneously false for all possible signs $\boldsymbol{\varepsilon}$. We refer the reader to Appendix E for further details.

3) OMP necessary condition including reachability assumptions: Our necessary condition for OMP is somewhat tricky because we must assume that $\mathcal{Q}$ is reachable by OMP using some input in $\mathrm{span}(\boldsymbol{A}_\mathcal{Q})$.

Theorem 5 ***[Necessary condition for OMP after q iterations]*** *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank and $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ is reachable. If ERC-OMP($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) does not hold, then there exists $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ for which OMP selects $\mathcal{Q}$ in the first q iterations and then a wrong atom $j \notin \mathcal{Q}^\star$ at iteration $(q+1)$.*

Theorem 5 is proved together with Theorem 4 in Appendix A-B. Setting aside the reachability issues, the principle of the proof is common to OMP and OLS. We proceed the proof of the sufficient condition (Theorem 3) backwards, as was done in [1, Theorem 3.10] in the case $\mathcal{Q} = \emptyset$.

In the special case where $q = 1$, Theorem 5 simplifies to a corollary similar to the OLS necessary condition (Theorem 4) because any subset $\mathcal{Q}$ of cardinality 1 is obviously reachable using the atom indexed by $\mathcal{Q}$ as input vector.

Corollary 2 ***[Necessary condition for OMP in the second iteration]*** *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank and let $i \in \mathcal{Q}^\star$. If ERC-OMP($\boldsymbol{A}, \mathcal{Q}^\star, \{i\}$) does not hold, then there exists $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ for which OMP selects $\boldsymbol{a}_i$ and then a wrong atom in the first two iterations.*

4) Discrimination between OMP and OLS at the k-th iteration: We provide an element of discrimination between OMP and OLS when their first $k-1$ iterations have selected true atoms, so that there is one remaining true atom which has not been chosen.

Theorem 6 ***[Guaranteed success of the k-th iteration of OLS]*** *If $[\boldsymbol{A}_{\mathcal{Q}^\star}, \boldsymbol{a}_j]$ is full rank for any $j \notin \mathcal{Q}^\star$, then ERC-OLS($\boldsymbol{A}, \mathcal{Q}^\star, k-1$) is true. Thus, if the first $k-1$ iterations of OLS select true atoms, the last true atom is necessarily selected in the k-th iteration.*

This result is straightforward from the definition of OLS in the optimization viewpoint: "OLS selects the new atom yielding the least possible residual" and because in the k-th iteration, the last true atom yields a zero valued residual. Another (analytical) proof of Theorem 6, given below, is based on the definition of ERC-OLS($\boldsymbol{A}, \mathcal{Q}^\star, k-1$). It will enable us to understand why the statement of Theorem 6 is not valid for OMP.

Proof: Assume that OLS yields a subset $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ after $k-1$ iterations. Let $\boldsymbol{a}_{\mathrm{last}}$ denote the last true atom so that $\boldsymbol{A}_{\mathcal{Q}^\star} = [\boldsymbol{A}_{\mathcal{Q}}, \boldsymbol{a}_{\mathrm{last}}]$ up to some column permutation. Since $\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star \backslash \mathcal{Q}}$ reduces to $\tilde{\boldsymbol{b}}^{\mathcal{Q}}_{\mathrm{last}}$ which is of unit norm, the pseudo-inverse $\tilde{\boldsymbol{B}}^\dagger_{\mathcal{Q}^\star \backslash \mathcal{Q}}$ takes the form $\left[\tilde{\boldsymbol{b}}^{\mathcal{Q}}_{\mathrm{last}}\right]^t$. Finally, (7) simplifies to:

$$F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) = |\langle \tilde{\boldsymbol{b}}^{\mathcal{Q}}_{\mathrm{last}}, \tilde{\boldsymbol{b}}^{\mathcal{Q}}_j \rangle| \leqslant 1 \tag{10}$$

since both vectors in the inner product are either of unit norm or equal to $\mathbf{0}$. Apply Lemma 8 in Appendix B: since for $j \notin \mathcal{Q}^\star$, $[\boldsymbol{A}_{\mathcal{Q}^\star}, \boldsymbol{a}_j]$ is full rank, $\left[\tilde{\boldsymbol{b}}^{\mathcal{Q}}_{\mathrm{last}}, \tilde{\boldsymbol{b}}^{\mathcal{Q}}_j\right]$ is full rank, thus (10) is a strict inequality. ∎

Similar to the calculation in the proof above, we rewrite $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$ defined in (6):

$$F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) = \frac{|\langle \tilde{\boldsymbol{a}}^{\mathcal{Q}}_{\mathrm{last}}, \tilde{\boldsymbol{a}}^{\mathcal{Q}}_j \rangle|}{\|\tilde{\boldsymbol{a}}^{\mathcal{Q}}_{\mathrm{last}}\|^2}. \tag{11}$$

However, we cannot ensure that $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) \leqslant 1$ since $\tilde{\boldsymbol{a}}^{\mathcal{Q}}_j$ and $\tilde{\boldsymbol{a}}^{\mathcal{Q}}_{\mathrm{last}}$ are not unit norm vectors. We refer the reader to subsection IV-C for a simple example with four atoms and two true atoms in which OMP is not guaranteed to select the second true atom when the first has already been chosen.

To further distinguish OMP and OLS, we elaborate a "bad recovery condition" under which OMP is guaranteed to fail in the sense that $\mathcal{Q}^\star$ is not reachable.

Theorem 7 ***[Sufficient condition for bad recovery with OMP]*** *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. If*

$$\min_{\substack{\mathcal{Q} \subsetneq \mathcal{Q}^\star \\ \mathrm{Card}[\mathcal{Q}]=k-1}} \left[\max_{j \notin \mathcal{Q}^\star} F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)\right] \geqslant 1, \qquad \text{BRC-OMP}(\boldsymbol{A}, \mathcal{Q}^\star)$$

then $\mathcal{Q}^\star$ cannot be reached by OMP using any input in $\mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$.

Specifically, BRC-OMP($\boldsymbol{A}, \mathcal{Q}^\star$) guarantees that a wrong selection occurs at the k-th iteration when the previous iterations have succeeded.

Proof: Assume that for some $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$, the first $k-1$ iterations of OMP succeed, *i.e.,* they select $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ of cardinality $k-1$. Let $\boldsymbol{a}_{\mathrm{last}}$ denote the last true atom ($\boldsymbol{A}_{\mathcal{Q}^\star} = [\boldsymbol{A}_{\mathcal{Q}}, \boldsymbol{a}_{\mathrm{last}}]$ up to some permutation of columns). The residual $\boldsymbol{r}_{\mathcal{Q}}$ yielded by OMP after $k-1$ iterations is obviously proportional to $\tilde{\boldsymbol{a}}^{\mathcal{Q}}_{\mathrm{last}}$.

BRC-OMP($\boldsymbol{A}, \mathcal{Q}^\star$) implies that ERC-OMP($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) is false, thus there exists $j \notin \mathcal{Q}^\star$ such that $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) \geqslant 1$. According to (11), $|\langle \tilde{\boldsymbol{a}}^{\mathcal{Q}}_{\mathrm{last}}, \tilde{\boldsymbol{a}}^{\mathcal{Q}}_j \rangle| \geqslant \|\tilde{\boldsymbol{a}}^{\mathcal{Q}}_{\mathrm{last}}\|^2$ thus $|\langle \boldsymbol{r}_{\mathcal{Q}}, \tilde{\boldsymbol{a}}^{\mathcal{Q}}_j \rangle| \geqslant |\langle \boldsymbol{r}_{\mathcal{Q}}, \tilde{\boldsymbol{a}}^{\mathcal{Q}}_{\mathrm{last}} \rangle|$. We conclude that $\boldsymbol{a}_{\mathrm{last}}$ cannot be chosen in the k-th iteration of OMP. ∎

Although BRC-OMP($\boldsymbol{A}, \mathcal{Q}^\star$) may appear restrictive (as a minimum is involved in the left-hand side), we will see in Section IV that it may be frequently met, especially when the atoms of $\boldsymbol{A}$ are strongly correlated.

IV. Empirical evaluation of the OMP and OLS recovery conditions

The purpose of this section is twofold. In subsection IV-B, we evaluate and compare the ERC-OMP and ERC-OLS conditions for several kinds of dictionaries. In particular, we study the dependence of $F^{\mathrm{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}} \triangleq \max_{j \notin \mathcal{Q}^\star} F^{\mathrm{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$ with respect to the dimensions m, n of the dictionary and the subset cardinalities $k = \mathrm{Card}\,[\mathcal{Q}^\star]$ and $q = \mathrm{Card}\,[\mathcal{Q}]$. This allows us to analyze, for random and deterministic dictionaries, from which iteration q the ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) condition may be met, *i.e.,* $F^{\mathrm{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}} < 1$. In subsection IV-C, we emphasize the distinction between OMP and OLS by showing that the bad recovery condition for OMP may be frequently met, especially when some dictionary atoms are strongly correlated.

A. Dictionaries under consideration

Our recovery conditions will be evaluated for three kinds of dictionaries.

We consider first randomly Gaussian dictionaries whose entries obey the standard Gaussian distribution. Once the dictionary elements are drawn, we normalize each atom in such a way that $\|\boldsymbol{a}_i\| = 1$.

"Hybrid" dictionaries are also studied, whose atoms result from an additive mixture of a deterministic (constant) and a random component. Specifically, we set $\boldsymbol{a}_i = \alpha_i(\boldsymbol{g}_i + t_i\mathbf{1})$ where $\boldsymbol{g}_i$ is drawn according to the standard Gaussian distribution, $\mathbf{1}$ is the (deterministic) vector whose entries are all equal to 1, and the scalar t_i obeys the uniform distribution on $[0, T]$, with $T > 0$. Once $\boldsymbol{g}_i$ and t_i are drawn, α_i is set in such a way that $\|\boldsymbol{a}_i\| = 1$. In this simulation, the mutual coherence is increased in comparison to the case $T = 0$ (*i.e.,* for randomly Gaussian dictionaries). The random vector $\boldsymbol{g}_i$ plays the role of a noise process added to the deterministic signal $t_i\mathbf{1}$. When T is large, the atom normalization makes the noise level very low in comparison with the deterministic component, thus the atoms are almost deterministic, and look alike each other.

Finally, we consider a sparse spike train deconvolution problem of the form $\boldsymbol{y} = \boldsymbol{h} * \boldsymbol{x}$, where $\boldsymbol{h}$ is a Gaussian impulse response of variance σ^2 (for simplicity, the smallest values in $\boldsymbol{h}$ are thresholded so that $\boldsymbol{h}$ has a finite support of width $\lceil 6\sigma \rceil$). This is a typical inverse problem in which the dictionary coherence is large. This problem is known to be a challenging one since both OMP and OLS are likely to yield false support recovery in practice [24–26]. This is also true for basis pursuit [27]. The problem can be reformulated as $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ where the dictionary $\boldsymbol{A}$ gathers shifted versions of the impulse response $\boldsymbol{h}$. To be more specific, we first consider a convolution operator with the same sampling rate for the input and output signals $\boldsymbol{x}$ and $\boldsymbol{y}$, and we set boundary conditions so that the convoluted signal $\boldsymbol{h} * \boldsymbol{x}$ resulting from $\boldsymbol{x}$ can be fully observed without truncation. Thus, $\boldsymbol{A}$ is a slightly undercomplete ($m > n$ with $m \approx n$) Toeplitz matrix. Alternately, we perform simulations in which the sampling rate of the input signal $\boldsymbol{x}$ is higher than that of $\boldsymbol{y}$ (*i.e.,* $\boldsymbol{y}$ results from a down-sampling of $\boldsymbol{h} * \boldsymbol{x}$), leading to an overcomplete dictionary $\boldsymbol{A}$ which does not have a Toeplitz structure anymore.

Regarding the last two problems, we found that the ERC factor $F_{\mathcal{Q}^\star} \triangleq F^{\mathrm{Oxx}}_{\mathcal{Q}^\star,\emptyset}$ which is the left hand-side in the ERC($\boldsymbol{A}, \mathcal{Q}^\star$) condition can become huge when T (respectively, σ) is increased. For instance, when T is equal to 10, 100 and 1000, the average value of $F_{\mathcal{Q}^\star}$ is equal to 7, 54 and 322, respectively, for a dictionary of size 100×1000 and for $k = 10$.

B. Evaluation of the ERC-Oxx conditions

We first show that for randomly Gaussian dictionaries, there is no systematic implication between the ERC-OMP($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) and ERC-OLS($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) conditions, nor between ERC-OMP($\boldsymbol{A}, \mathcal{Q}^\star, q$) and ERC-OLS($\boldsymbol{A}, \mathcal{Q}^\star, q$). Then, we perform more complete numerical simulations to assess the dependence of $F^{\mathrm{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}}$ with respect to the size (m, n) of the dictionary and the subset cardinalities k and q for the three kinds of dictionaries. We will build "phase transition diagrams" (in a sense to be defined below) to compare the OMP and OLS recovery conditions. The general principle of our simulations is 1) to draw the dictionary $\boldsymbol{A}$ and the subset $\mathcal{Q}^\star$; and 2) to gradually increase $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ by one element until ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) is met.

1) There is no logical implication between the ERC-OMP and ERC-OLS conditions: We first investigate what is going on after the first iteration ($q = 1$). We compare ERC-OMP($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) and ERC-OLS($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) for a common dictionary and given subsets $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ with $q = 1$. As the recovery conditions take the form "for all $j \notin \mathcal{Q}^\star$, $F^{\mathrm{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) < 1$", it is sufficient to just consider the case where there is one wrong atom $\boldsymbol{a}_j$ to study the logical implication between the ERC-OMP and ERC-OLS conditions. Therefore, in this paragraph, we consider undercomplete dictionaries $\boldsymbol{A}$ with $k + 1$ atoms. Testing ERC($\boldsymbol{A}, \mathcal{Q}^\star$), ERC-OMP($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) and ERC-OLS($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) amounts to evaluating $F_{\mathcal{Q}^\star}(\boldsymbol{a}_j)$, $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$ and $F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$ and comparing them to 1.

Fig. 1 is a scatter plot of the three criteria for 10.000 randomly Gaussian dictionaries $\boldsymbol{A}$ of size 100×11. The subset $\mathcal{Q} = \{1\}$ is systematically chosen as the first atom of $\boldsymbol{A}$. Figs. 1(a,b) are in good agreement with Lemma 2: we verify that $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) \leqslant F_{\mathcal{Q}^\star}(\boldsymbol{a}_j)$ whether the ERC holds or not, and that $F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) \leqslant F_{\mathcal{Q}^\star}(\boldsymbol{a}_j)$ systematically occurs only when $F_{\mathcal{Q}^\star}(\boldsymbol{a}_j) < 1$. On Fig. 1(c) displaying $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$ versus $F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$, we only keep the trials for which $F_{\mathcal{Q}^\star}(\boldsymbol{a}_j) \geqslant 1$, *i.e.,* ERC($\boldsymbol{A}, \mathcal{Q}^\star$) does not hold. Since both south-east and north-west quarter planes are populated, we conclude that neither OMP nor OLS is uniformly better. To be more specific, when ERC-OMP($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) holds but ERC-OLS($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) does not, there exists an input $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ for which OLS selects $\mathcal{Q} = \{1\}$ and then a wrong atom in the second iteration (Theorem 4). On the contrary, OMP is guaranteed to exactly recover this input according to Theorem 3. The same situation can occur when inverting the roles of OMP and OLS according to Corollary 2 and Theorem 3 (note that this analysis becomes more complex when $\mathrm{Card}\,[\mathcal{Q}] \geqslant 2$ since ERC-OMP($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) alone is not a necessary condition for OMP anymore; Theorem 5 also involves the assumption that $\mathcal{Q}$ is reachable).

We have compared ERC-OMP($\boldsymbol{A}, \mathcal{Q}^\star, 1$) and ERC-OLS($\boldsymbol{A}, \mathcal{Q}^\star, 1$), which take into account all the possible subsets of $\mathcal{Q}^\star$ of cardinality 1. Again, we found that when ERC($\boldsymbol{A}, \mathcal{Q}^\star$) is not met, it can occur that ERC-

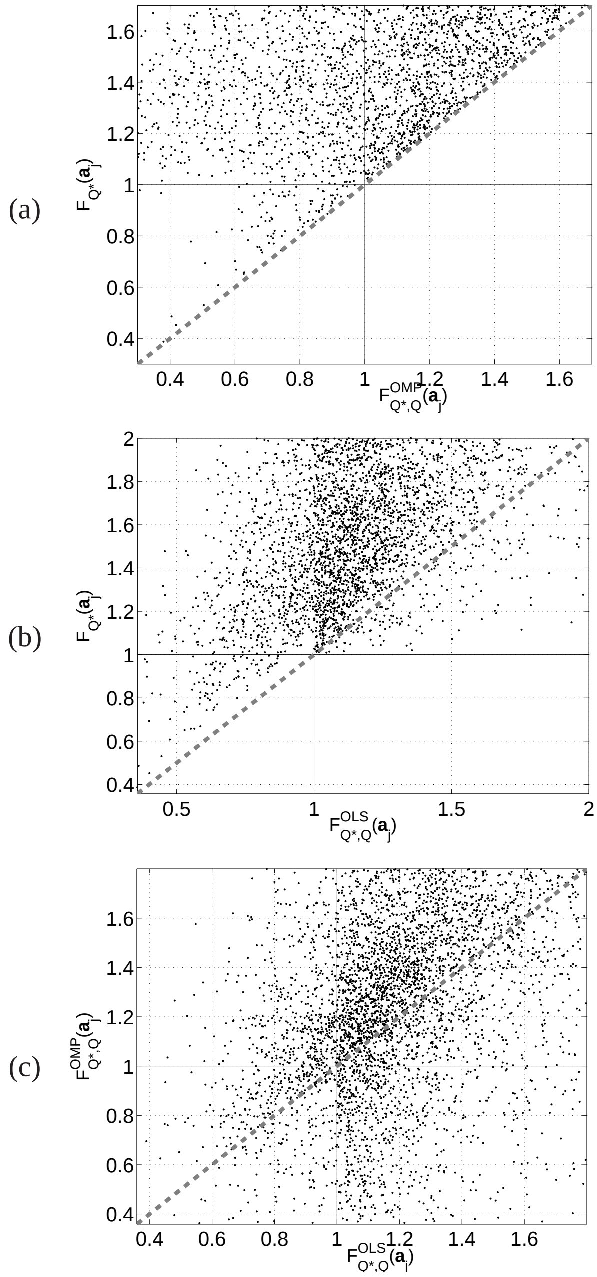
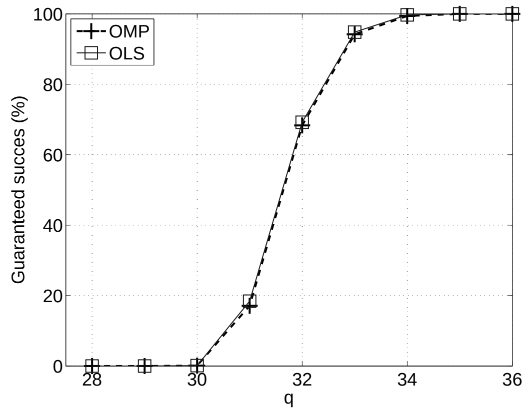


Fig. 1. Comparison of the OMP and OLS exact recovery conditions. We draw 10.000 Gaussian dictionaries of size 100×11 and set $k = 10$ so that there is only one wrong atom $\boldsymbol{a}_j$. $\mathcal{Q}$ is always set to the first atom ($\mathrm{Card}\,[\mathcal{Q}] = 1$). Plot of (a) $F_{\mathcal{Q}^\star}(\boldsymbol{a}_j)$ *vs* $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$; (b) $F_{\mathcal{Q}^\star}(\boldsymbol{a}_j)$ *vs* $F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$; (c) $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$ *vs* $F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$. For the last subfigure, we keep the trials for which $F_{\mathcal{Q}^\star}(\boldsymbol{a}_j) \geqslant 1$.

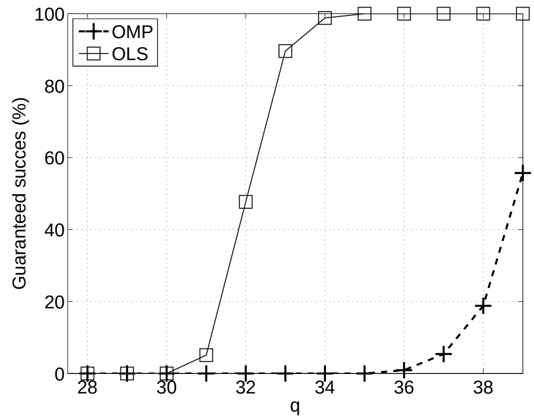
OMP($\boldsymbol{A}, \mathcal{Q}^\star, 1$) holds while ERC-OLS($\boldsymbol{A}, \mathcal{Q}^\star, 1$) does not and *vice versa*.

2) Phase transition analysis for overcomplete random dictionaries: We now address the case of overcomplete dictionaries. Moreover, we study the dependence of the ERC-Oxx conditions with respect to the cardinalities k and q for $k > q \geqslant 2$ and we compare them for common problems ($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$).

Let us start with simple preliminary remarks. Because the ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) conditions take the form "for all $j \notin \mathcal{Q}^\star$, $F^{\mathrm{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) < 1$", they are more often met when the dictionary is undercomplete (or when $m \approx n$) than in the overcomplete case: when the submatrix $\boldsymbol{A}_{\mathcal{Q}^\star}$ gathering the true atoms is given, $\max_{j \notin \mathcal{Q}^\star} F^{\mathrm{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$ is obviously increasing when additional wrong atoms $\boldsymbol{a}_j$ are incorporated, *i.e.,* when n is increasing. Additionally, we notice that for given $\boldsymbol{A}$ and $\mathcal{Q}^\star$, $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}$ always decreases when $\mathcal{Q}$ is growing by definition of $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}$. This might not be the case of $F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}$ for specific settings but it happens to be true in average for random dictionaries.



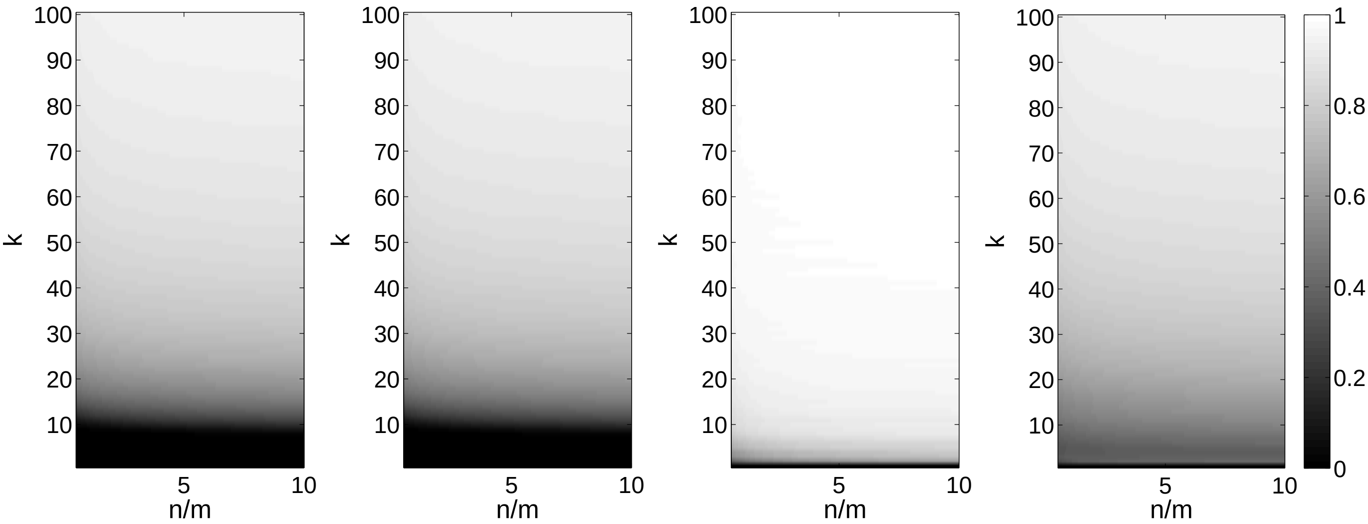
(a) Random dictionaries ($T = 0$)



(b) Hybrid dictionaries ($T = 10$)

Fig. 2. Phase transition curves: for each $q < k$, we count the rate of trials where ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) is true, with $\mathrm{Card}\,[\mathcal{Q}] = q$. The dictionaries are of size 200×600, k is set to 40 and 1,000 Monte Carlo trials are performed. (a) Randomly Gaussian dictionaries; (b) Hybrid dictionaries with $T = 10$.

In the following experiments, $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ is gradually increased for fixed $\boldsymbol{A}$ and $\mathcal{Q}^\star$, and we search for the first cardinality $q = \mathrm{Card}\,[\mathcal{Q}]$ for which ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) is met. This allows us to define a "phase transition curve" [17, 28] which separates the q-values for which ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) is never met, and is always met. Examples of phase transition curves are given on Fig. 2



(a) OMP, $T = 0$ (b) OLS, $T = 0$ (c) OMP, $T = 10$ (d) OLS, $T = 10$

Fig. 3. Phase transition diagrams for the ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) condition. The gray levels represent the ratio $[q]^{\mathrm{Oxx}}(m, n, k)/k \in [0, 1]$. Averaging is done over 200 draws of dictionary $\boldsymbol{A}$ and subset $\mathcal{Q}^\star$. (a,b) Randomly Gaussian dictionaries of size $200 \times n$ with $n \leqslant 2000$; (c,d) Hybrid dictionaries of same size, with $T = 10$.

for random ($T = 0$) and hybrid dictionaries ($T = 10$). Fig. 2(a) shows that for $T = 0$, the phase transition regime occurs in the same interval $q \in \{30, \dots, 34\}$ for both OMP and OLS and that the OMP and OLS curves are superimposed. On the contrary, for hybrid dictionaries (Fig. 2(b)), the mutual coherence increases and the OLS curve is significantly above the OMP curve. Thus, the guaranteed success for OLS occurs (in average) for an earlier iteration than for OMP. For larger values of T (*e.g.,* for $T = 100$), the ERC-OMP condition is never met before $q = k - 1$, and even for $q = k - 1$, it is met for only $4\ \%$ of trials.

The experiment of Fig. 2 is repeated for many values of k and dictionary sizes $m \times n$. For given $\boldsymbol{A}$ and $\mathcal{Q}^\star$, let $q^{\mathrm{Oxx}}(m, n, k)$ denote the lowest value of $q = \mathrm{Card}\,[\mathcal{Q}]$ for which ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) is true. For random and hybrid dictionaries, we perform 200 Monte Carlo simulations in which random matrices $\boldsymbol{A}$ and subsets $(\mathcal{Q}^\star, \mathcal{Q})$ are drawn and we compute the average values of q^{Oxx}, denoted by $[q]^{\mathrm{Oxx}}(m, n, k)$. This yields a phase transition diagram [12, 29] with the dictionary size (*e.g.,* n/m) and the sparsity level k in x- and y-axes, respectively. In this image, the gray levels represent the ratio $[q]^{\mathrm{Oxx}}(m, n, k)/k$ (see Fig. 3). Note that our phase transition diagrams are related to worst case recovery conditions, so better performance may be achieved by actually running Oxx for some simulated data $(\boldsymbol{y}, \boldsymbol{A})$ and testing whether the support $\mathcal{Q}^\star$ is found, where $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}^\star$ and the unknown nonzero amplitudes in $\boldsymbol{x}^\star$ are drawn according to a specific distribution.

A general comment regarding the results of Fig. 3 is that the ERC-Oxx conditions are satisfied early (for low values of q/k) when the unknown signal is highly sparse (k is low) or when n/m is low, *i.e.,* when the dictionary is not highly overcomplete. The ratio $[q]^{\mathrm{Oxx}}(m, n, k)/k$ gradually grows with k and n/m. Regarding the OMP *vs* OLS comparison, the phase diagrams obtained for OMP and OLS look very much alike for Gaussian dictionaries ($T = 0$). On the contrary, we observe drastic differences in favor of OLS for hybrid dictionaries (Fig. 3(c,d)): $F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}$ is significantly lower than $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}$.

We have performed similar tests for randomly uniform dictionaries (and hybrid dictionaries based on a randomly uniform process) and we draw conclusions similar to the Gaussian case. We have not encountered any situation where $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}$ is (in average) significantly lower than $F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}$.

3) ERC-Oxx evaluation for sparse spike train deconvolution dictionaries: We reproduced the above experiments for the convolutive dictionary introduced in subsection IV-A. Since the dictionary is deterministic, only one trial is performed per cardinality (m, n, k). In each of the simulations hereafter, we set $\mathcal{Q}$ and $\mathcal{Q}^\star$ to contiguous atoms. This is the worst situation because contiguous atoms are the most highly correlated and exact support recovery may be more easily achieved if we impose a minimum distance between true atoms [24, 30]. The curves of Fig. 4 represent $F^{\mathrm{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}}$ with respect to q for some given $(\boldsymbol{A}, \mathcal{Q}^\star)$. It is noticeable that the OLS curve decreases much faster than the OMP curve, and that

(a)

(b)

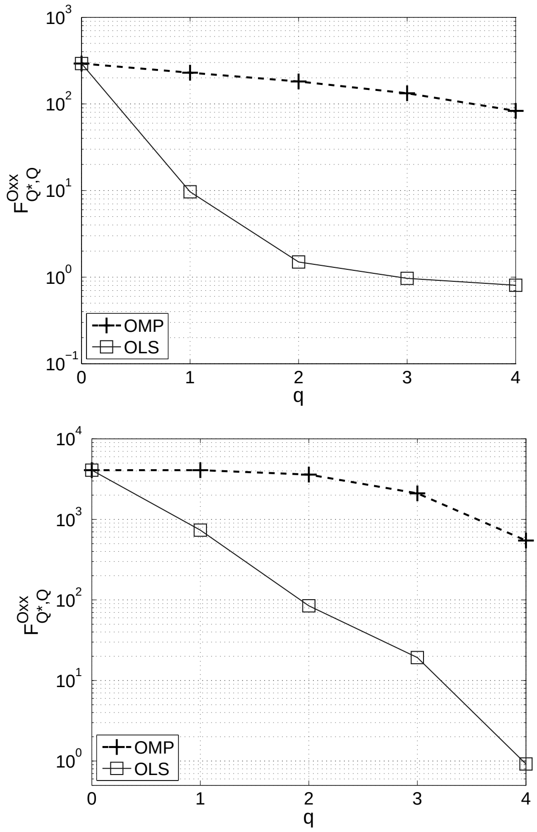


Fig. 4. Curve representing $F^{\mathrm{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}}$ as a function of $q = \mathrm{Card}\,[\mathcal{Q}]$ for the Gaussian deconvolution problem. $\mathcal{Q}^\star$ is fixed and $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ is gradually growing. $\mathcal{Q}^\star$ and $\mathcal{Q}$ are formed of the first $k = 5$ and the first q atoms, respectively with $q < k$. (a) The Gaussian impulse response is of width $\sigma = 50$ and the dictionary is of size 3000×2710. (b) σ is set to 10, and the dictionary is of size 1000×4940.

$F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}$ remains huge even after a number of iterations. For all our trials where the true atoms strongly overlap, the ERC-OMP($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) condition is not met while ERC-OLS($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) may be fulfilled after a number of iterations which is, however, close to k. Moreover, we found that when σ is large enough, $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}$ remains larger than 1 even for $q = k-1$, whereas the ERC-OLS condition is always met for $q = k-1$.

Empirical evaluations of the ERC condition for sparse spike train deconvolution was already done in [27]. In [24, 27, 30], a stronger sufficient condition than the ERC is evaluated for convolutive dictionaries. It is a sufficient (but not necessary) exact recovery condition that is easier to compute than the ERC because it does not require any matrix inversion, and only relies on inner products between the dictionary atoms (see [31, Lemma 3] for further details). In [27, 30], it was pointed out that the ERC condition is usually not fulfilled for convolutive dictionaries, but when the true atoms are enough spaced, successful recovery is guaranteed to occur. Our study can be seen as an alternative analysis to [24, 27, 30] in which no minimal distance constraint is imposed.

C. Examples where the bad recovery condition of OMP is met

We exhibit several situations in which the BRC-OMP($\boldsymbol{A}, \mathcal{Q}^\star$) condition may be fulfilled. This allows us to distinguish OMP from OLS as we know that under regular conditions, any subset $\mathcal{Q}^\star$ is reachable using OLS at least for some input in $\mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ (Lemma 3). The first situation is a simple dictionary with four atoms, some of which being strongly correlated. For this example, we show a stronger result than the BRC: there exists a subset $\mathcal{Q}^\star$ which is not reachable for any $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$, but not even for any $\boldsymbol{y} \in \mathbb{R}^m$. The other examples involve the random, hybrid and deterministic dictionaries introduced in subsection IV-A.

1) Example with four atoms:

Example 1 *Consider the simple dictionary*

$$\boldsymbol{A} = \begin{bmatrix} \cos\theta_1 & \cos\theta_1 & 0 & 0 \\ -\sin\theta_1 & \sin\theta_1 & \cos\theta_2 & \cos\theta_2 \\ 0 & 0 & \sin\theta_2 & -\sin\theta_2 \end{bmatrix}$$

with $\mathcal{Q}^\star = \{1, 2\}$. Set θ_2 to an arbitrary value in $(0, \pi/2)$. When $\theta_1 \neq 0$ is close enough to 0, BRC-OMP($\boldsymbol{A}, \mathcal{Q}^\star$) is met. Moreover, OMP cannot reach $\mathcal{Q}^\star$ in two iterations for any $\boldsymbol{y} \in \mathbb{R}^3$ (specifically, when $\boldsymbol{y} \in \mathbb{R}^3$ is proportional to neither $\boldsymbol{a}_1$ nor $\boldsymbol{a}_2$, $\boldsymbol{a}_3$ or $\boldsymbol{a}_4$ is selected in the first two iterations).

Proof of Example 1: We first prove that the BRC condition is met by calculating the factors $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\{1\}}(\boldsymbol{a}_j)$ and $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\{2\}}(\boldsymbol{a}_j)$ for $j \in \{3, 4\}$. Let us start with $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\{1\}}(\boldsymbol{a}_j)$.

The simple projection calculation $\tilde{\boldsymbol{a}}_i = \boldsymbol{a}_i - \langle \boldsymbol{a}_i, \boldsymbol{a}_1 \rangle \boldsymbol{a}_1$ (the tilde notation implicitly refers to $\mathcal{Q} = \{1\}$) leads to:

$$\tilde{\boldsymbol{a}}_2 = \sin(2\theta_1) \begin{bmatrix} \sin\theta_1 \\ \cos\theta_1 \\ 0 \end{bmatrix}, \quad \tilde{\boldsymbol{a}}_3 = \begin{bmatrix} \sin\theta_1 \cos\theta_1 \cos\theta_2 \\ \cos^2\theta_1 \cos\theta_2 \\ \sin\theta_2 \end{bmatrix}$$
$$\text{and } \tilde{\boldsymbol{a}}_4 = \begin{bmatrix} \sin\theta_1 \cos\theta_1 \cos\theta_2 \\ \cos^2\theta_1 \cos\theta_2 \\ -\sin\theta_2 \end{bmatrix}.$$

According to (11), the OMP recovery factor reads for $j \in \{3, 4\}$:

$$F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\{1\}}(\boldsymbol{a}_j) = \frac{|\langle \tilde{\boldsymbol{a}}_2, \tilde{\boldsymbol{a}}_j \rangle|}{\|\tilde{\boldsymbol{a}}_2\|^2} = \frac{|\cos\theta_1 \cos\theta_2|}{|\sin(2\theta_1)|} \tag{12}$$

given that $\|\tilde{\boldsymbol{a}}_2\| = |\sin(2\theta_1)|$ and $|\langle \tilde{\boldsymbol{a}}_2, \tilde{\boldsymbol{a}}_3 \rangle| = |\langle \tilde{\boldsymbol{a}}_2, \tilde{\boldsymbol{a}}_4 \rangle| = \|\tilde{\boldsymbol{a}}_2\| \, |\cos\theta_1 \cos\theta_2|$. $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\{2\}}(\boldsymbol{a}_j)$ can be obtained symmetrically by replacing θ_1 by $-\theta_1$ in (12).

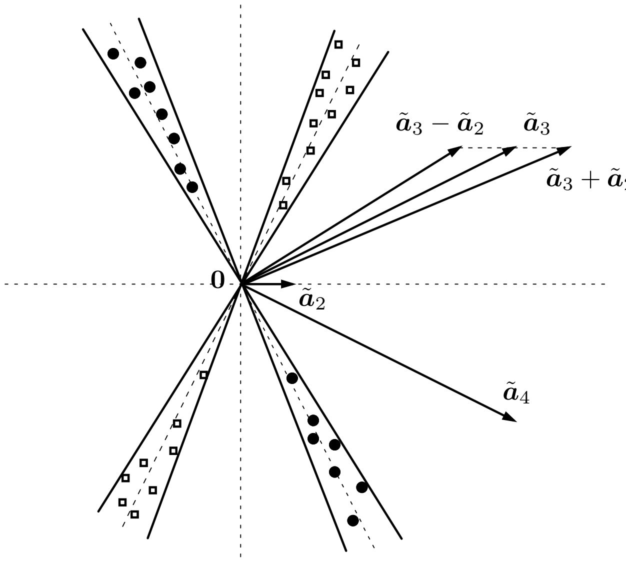


Fig. 5. Example 1: drawing of the plane $\mathrm{span}(\boldsymbol{a}_1)^\perp$. The tilde notation refers to the subset $\mathcal{Q} = \{1\}$. When θ_1 is close to 0, $\tilde{\boldsymbol{a}}_2$ is of very small norm since $\boldsymbol{a}_2$ is almost equal to $\boldsymbol{a}_1$, while $\boldsymbol{a}_3$ and $\boldsymbol{a}_4$, which are almost orthogonal to $\boldsymbol{a}_1$, yield projections $\tilde{\boldsymbol{a}}_3$ and $\tilde{\boldsymbol{a}}_4$ that are almost of unit norm. The angles $(\tilde{\boldsymbol{a}}_2, \tilde{\boldsymbol{a}}_3)$ and $(\tilde{\boldsymbol{a}}_2, \tilde{\boldsymbol{a}}_4)$ tend to θ_2 and $-\theta_2$ when $\theta_1 \to 0$. The bullet and square points correspond to positions $\boldsymbol{r}$ satisfying $|\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_2\rangle| \geqslant |\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_3\rangle|$ and $|\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_2\rangle| \geqslant |\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_4\rangle|$, respectively. The central directions of these two cones are orthogonal to $\tilde{\boldsymbol{a}}_3$ and $\tilde{\boldsymbol{a}}_4$, respectively (dashed lines). Both cones only intersect at $\boldsymbol{r} = \mathbf{0}$, therefore OMP cannot successively select $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ in the first two iterations.

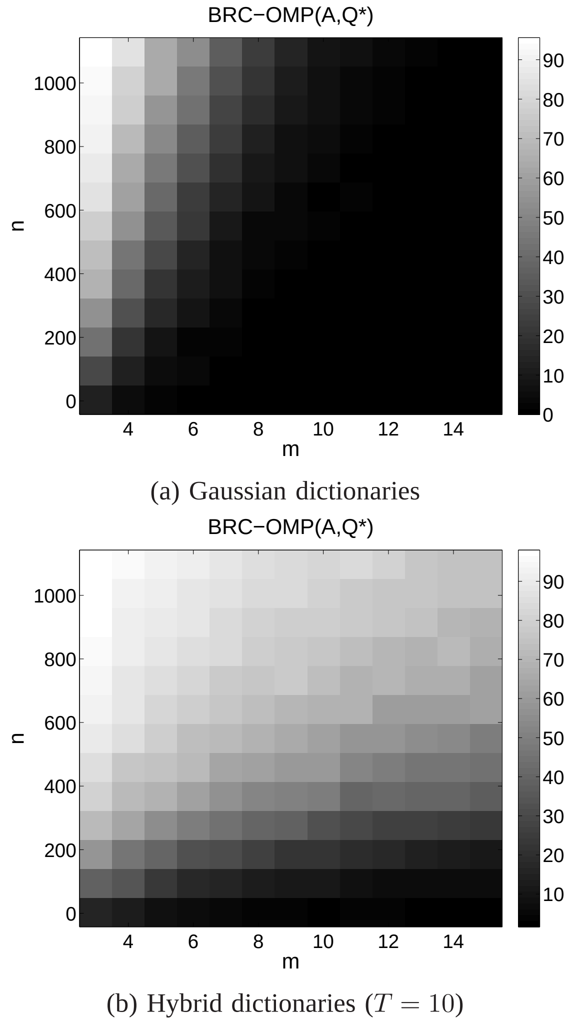


Fig. 6. Evaluation of the bad recovery condition BRC-OMP($\boldsymbol{A}, \mathcal{Q}^\star$) for randomly Gaussian (a) and hybrid (b) dictionaries of various sizes (m, n). 1,000 trials are performed per dictionary size, and $\mathcal{Q}^\star$ is always set to the first two atoms ($k = 2$). The gray levels correspond to the rate of guaranteed failure, *i.e.*, the proportion of trials where BRC-OMP($\boldsymbol{A}, \mathcal{Q}^\star$) holds.

Thus, we have $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\{2\}}(\boldsymbol{a}_j) = F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\{1\}}(\boldsymbol{a}_j)$. It follows that the left hand-side of the BRC-OMP($\boldsymbol{A}, \mathcal{Q}^\star$) condition reads (12) and tends towards $+\infty$ when θ_1 tends towards 0. Therefore, BRC-OMP($\boldsymbol{A}, \mathcal{Q}^\star$) is met when $|\theta_1|$ is small enough.

To show that $\mathcal{Q}^\star$ is not reachable for any $\boldsymbol{y} \in \mathbb{R}^3$, let us assume that OMP selects a true atom in the first iteration. Because there is a symmetry between $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$, we can assume without loss of generality that $\boldsymbol{a}_1$ is selected. Then, the data residual $\boldsymbol{r}$ after the first iteration lies in $\mathrm{span}(\boldsymbol{a}_1)^\perp$ which is of dimension 2. We show using geometrical arguments, that $\boldsymbol{a}_2$ cannot be selected in the second iteration for any $\boldsymbol{r} \in \mathrm{span}(\boldsymbol{a}_1)^\perp \backslash \{\mathbf{0}\}$. We refer the reader to Fig. 5 for a 2D display of the projected atoms in the plane $\mathrm{span}(\boldsymbol{a}_1)^\perp$.

Let $\mathcal{C}$ denote the set of points $\boldsymbol{r} \in \mathbb{R}^2$ satisfying $|\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_2\rangle| \geqslant |\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_3\rangle|$. $\boldsymbol{r} \in \mathcal{C}$ *if and only if* there exist $(\varepsilon_2, \varepsilon_3) \in \{-1, 1\}^2$ such that $\varepsilon_2\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_2\rangle \geqslant \varepsilon_3\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_3\rangle \geqslant 0$, *i.e.,*

$$\langle \boldsymbol{r}, \varepsilon_2\tilde{\boldsymbol{a}}_2 - \varepsilon_3\tilde{\boldsymbol{a}}_3\rangle \geqslant 0 \ \text{ and } \ \langle \boldsymbol{r}, \varepsilon_3\tilde{\boldsymbol{a}}_3\rangle \geqslant 0. \tag{13}$$

For each sign pattern $(\varepsilon_2, \varepsilon_3)$, (13) yields a 2D half cone defined as the intersection of two half-planes delimited by the directions which are orthogonal to $\tilde{\boldsymbol{a}}_3$ and $\varepsilon_2\tilde{\boldsymbol{a}}_2 - \varepsilon_3\tilde{\boldsymbol{a}}_3$. Moreover, the opposite sign pattern $(-\varepsilon_2, -\varepsilon_3)$ yields the remaining part of the same 2D cone. Consequently, the four possible sign patterns $(\varepsilon_2, \varepsilon_3) \in \{-1, 1\}^2$ yield both cones delimited by the orthogonal directions to $\tilde{\boldsymbol{a}}_3$ and $\tilde{\boldsymbol{a}}_2 + \tilde{\boldsymbol{a}}_3$, and to $\tilde{\boldsymbol{a}}_3$ and $-\tilde{\boldsymbol{a}}_2 + \tilde{\boldsymbol{a}}_3$, respectively. Because these cones are adjacent, their union $\mathcal{C}$ is the cone delimited by the orthogonal directions to $\tilde{\boldsymbol{a}}_3 + \tilde{\boldsymbol{a}}_2$ and $\tilde{\boldsymbol{a}}_3 - \tilde{\boldsymbol{a}}_2$ (plain lines in the south-east and north-west directions in Fig. 5). Similarly, the condition $|\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_2\rangle| \geqslant |\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_4\rangle|$ yields another 2D cone whose central direction is orthogonal to $\tilde{\boldsymbol{a}}_4$. When θ_1 is close to 0, both cones only intersect at $\boldsymbol{r} = \mathbf{0}$ (since their inner angle tends towards 0), thus

$$\forall \boldsymbol{r} \in \mathbb{R}^2 \backslash \{\mathbf{0}\}, \ |\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_2\rangle| < \max(|\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_3\rangle|, |\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_4\rangle|).$$

We conclude that $\boldsymbol{a}_2$ cannot be selected in the second iteration according to the OMP rule (1). ■

2) Numerical simulation of the BRC condition: We test the BRC-OMP condition for various dictionary sizes (m, n) for the random, hybrid and convolutive dictionaries introduced in subsection IV-A. The average results related to the random and hybrid dictionaries are gathered in Fig. 6 in the case $k = 2$. For randomly Gaussian dictionaries, we observe that the BRC-OMP condition may be met for strongly overcomplete dictionaries, *i.e.*, when $n \gg m$ (Fig. 6 (a)). In the special case $k = 2$, it is noticeable that OLS performs at least as well as OMP whether the BRC condition if fulfilled or not: when the first iteration (common to both algorithms) has succeeded, OLS cannot fail according to Theorem 6 while OMP is guaranteed to fail in cases where the BRC holds. For the hybrid dictionaries, the BRC condition is more frequently met when the dictionary is moderately overcomplete, *i.e.*, for large values of m/n. This result is in coherence with our evaluations of the ERC-Oxx condition (see, *e.g.*, Fig. 3(c)) which are more rarely met for hybrid dictionary than for random dictionaries.

We performed similar tests for the sparse spike train deconvolution problem with a Gaussian impulse response of width σ, and with $k = 2$ (the true atoms are contiguous, thus they are strongly correlated). We repeated the simulation of Fig. 6 for various sizes $m \approx n$ and various widths σ, and we found that whatever (m, n), the BRC condition is always met for $\sigma \geqslant 1.5$ and never met when $\sigma \leqslant 1.4$. The images of Fig. 6 thus become uniformly white and uniformly black, respectively. To be more specific, the value of the left hand-side of the BRC-OMP($\boldsymbol{A}, \mathcal{Q}^\star$) condition gradually increases with σ, *e.g.*, this value reaches 10, 35 and 48 for $\sigma = 10$, 20 and 50, respectively for dictionaries of size $m \approx n$, with $m = 3000$. This result is in coherence with that of Fig. 4 which already indicated that the $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}$ factor becomes huge for convolutive problems with strongly correlated atoms.

Note that when $\mathcal{Q}^\star$ does not involve contiguous atoms but "spaced atoms" which are less correlated, the bad recovery condition are met for larger values of σ: denoting by Δ the minimum distance between two true atoms, the lowest Δ value for which the BRC is met turns out to be an increasing affine function of σ. Similar empirical studies were done in [27] for the exact recovery condition for spaced atoms, and in [24, 27] for the weak exact recovery condition of [31, Lemma 3]. In particular, the numerical simulations in [24] for the Gaussian deconvolution problem demonstrate that the latter condition is met for larger σ's when the minimum distance between true atoms is increased and the limit Δ value corresponding to the phase transition is also an affine function of σ. Our bad recovery condition results are thus a complement to those of [24].

V. Conclusions

Our first contribution is an original analysis of OLS based on the extension of the ERC condition. We showed that when the ERC holds, OLS is guaranteed to yield an exact support recovery. Although OLS has been acknowledged in several communities for two decades, such a theoretical analysis was lacking. Our second contribution is a parallel study of OMP and OLS when a number of iterations have been performed and true atoms have been selected. We found that neither OMP nor OLS is uniformly better. In particular, we showed using randomly Gaussian dictionaries that when the ERC is not met but the first iteration (which is common to OMP and OLS) selects a true atom, there are counter-examples for which OMP is guaranteed to yield an exact support recovery while OLS does not, and *vice versa*.

Finally, several elements of analysis suggest that OLS behaves better than OMP. First, any subset $\mathcal{Q}$ can be reached by OLS using some input in $\mathrm{span}(\boldsymbol{A}_{\mathcal{Q}})$ while for some dictionaries, it may occur that some subsets are never reached by OMP for any $\boldsymbol{y} \in \mathbb{R}^m$. In other words, OLS has a stronger capability of exploration. Secondly, when all true atoms except one have been found by OLS and no wrong selection occurred, OLS is guaranteed to find the last true atom in the following iteration while OMP may fail.

For problems in which the dictionary is far from orthogonal and some dictionary atoms are strongly correlated, we found in our experiments that the OLS recovery condition might be met after some iterations while the OMP recovery condition is rarely met. We did not encounter the opposite situation where the OMP recovery condition is frequently met after fewer iterations than the OLS condition. Moreover, guaranteed failure of OMP may occur more often when the dictionary coherence is large. These results are in coherence with empirical studies reporting that OLS usually outperforms OMP at the price of a larger numerical cost [9, 11]. In our experience, OLS yields a residual error which may be by far lower than that of OMP after the same number of iterations [25]. Moreover, it performs better support recoveries in terms of ratio between the number of good detections and of false alarms [26].

Appendix A
Necessary and sufficient conditions of exact recovery for OMP and OLS

This appendix includes the complete analysis of our OMP and OLS recovery conditions.

$$F^{\text{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) = F^{\text{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}'}(\boldsymbol{a}_j) + \left|\left(\boldsymbol{A}^\dagger_{\mathcal{Q}^\star}\boldsymbol{a}_j\right)(\ell)\right| \tag{20}$$

$$F^{\text{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) = \left|\chi^{\mathcal{Q},\mathcal{Q}'}_j - \eta^{\mathcal{Q},\mathcal{Q}'}_j \sum_{i\in\mathcal{Q}^\star\backslash\mathcal{Q}'} \frac{\boldsymbol{\beta}^{\mathcal{Q}^\star\backslash\mathcal{Q}'}_j(i)\chi^{\mathcal{Q},\mathcal{Q}'}_i}{\eta^{\mathcal{Q},\mathcal{Q}'}_i}\right| + \eta^{\mathcal{Q},\mathcal{Q}'}_j \sum_{i\in\mathcal{Q}^\star\backslash\mathcal{Q}'} \frac{\left|\boldsymbol{\beta}^{\mathcal{Q}^\star\backslash\mathcal{Q}'}_j(i)\right|}{\eta^{\mathcal{Q},\mathcal{Q}'}_i} \tag{21}$$

A. Sufficient conditions

We show that when Oxx happens to select true atoms during its early iterations, it is guaranteed to recover the whole unknown support in the subsequent iterations when the ERC-Oxx($\boldsymbol{A},\mathcal{Q}^\star,\mathcal{Q}$) condition is fulfilled. We establish Theorem 3 whose direct consequence is Theorem 2 stating that when ERC($\boldsymbol{A},\mathcal{Q}^\star$) holds, OLS is guaranteed to succeed.

1) ERC-Oxx are sufficient recovery conditions at a given iteration: We follow the analysis of [1, Theorem 3.1] to extend Tropp's exact recovery condition to a sufficient condition dedicated to the $(q+1)$-th iteration of Oxx.

Lemma 4 *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. If Oxx with $\boldsymbol{y}\in\operatorname{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ as input selects q true atoms $\mathcal{Q}\subsetneq\mathcal{Q}^\star$ and ERC-Oxx($\boldsymbol{A},\mathcal{Q}^\star,\mathcal{Q}$) holds, then the $(q+1)$-th iteration of Oxx selects a true atom.*

Proof: According to the selection rule (1)-(2), Oxx selects a true atom at iteration $(q+1)$ *if and only if*

$$\phi(\boldsymbol{r}_\mathcal{Q}) \triangleq \frac{\max_{i\notin\mathcal{Q}^\star}|\langle\boldsymbol{r}_\mathcal{Q},\tilde{\boldsymbol{c}}_i\rangle|}{\max_{i\in\mathcal{Q}^\star\backslash\mathcal{Q}}|\langle\boldsymbol{r}_\mathcal{Q},\tilde{\boldsymbol{c}}_i\rangle|} < 1. \tag{14}$$

Let us gather the vectors $\tilde{\boldsymbol{c}}_i$ indexed by $i\notin\mathcal{Q}^\star$ and $i\in\mathcal{Q}^\star\backslash\mathcal{Q}$ in two matrices $\tilde{\boldsymbol{C}}_{\bullet\backslash\mathcal{Q}^\star}$ and $\tilde{\boldsymbol{C}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}$ of dimensions $m\times(n-k)$ and $m\times(k-q)$, respectively where the notation $\bullet$ stands for all indices $i\in\{1,\ldots,n\}$. The condition (14) rereads:

$$\phi(\boldsymbol{r}_\mathcal{Q}) = \frac{\|\tilde{\boldsymbol{C}}^t_{\bullet\backslash\mathcal{Q}^\star}\boldsymbol{r}_\mathcal{Q}\|_\infty}{\|\tilde{\boldsymbol{C}}^t_{\mathcal{Q}^\star\backslash\mathcal{Q}}\boldsymbol{r}_\mathcal{Q}\|_\infty} < 1.$$

Following Tropp's analysis, we re-arrange the vector $\boldsymbol{r}_\mathcal{Q}$ occurring in the numerator. Since $\boldsymbol{r}_\mathcal{Q} = \boldsymbol{P}^\perp_\mathcal{Q}\boldsymbol{y}$ and $\boldsymbol{y}\in\operatorname{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$, $\boldsymbol{r}_\mathcal{Q}\in\operatorname{span}(\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}) = \operatorname{span}(\tilde{\boldsymbol{C}}_{\mathcal{Q}^\star\backslash\mathcal{Q}})$. We rewrite $\boldsymbol{r}_\mathcal{Q}$ as $\tilde{\boldsymbol{P}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}\boldsymbol{r}_\mathcal{Q}$ where $\tilde{\boldsymbol{P}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}$ stands for the orthogonal projector on $\operatorname{span}(\tilde{\boldsymbol{C}}_{\mathcal{Q}^\star\backslash\mathcal{Q}})$: $\tilde{\boldsymbol{P}}_{\mathcal{Q}^\star\backslash\mathcal{Q}} = \tilde{\boldsymbol{P}}^t_{\mathcal{Q}^\star\backslash\mathcal{Q}} = \left(\tilde{\boldsymbol{C}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}\tilde{\boldsymbol{C}}^\dagger_{\mathcal{Q}^\star\backslash\mathcal{Q}}\right)^t$. $\phi(\boldsymbol{r}_\mathcal{Q})$ rereads

$$\phi(\boldsymbol{r}_\mathcal{Q}) = \frac{\|\left(\tilde{\boldsymbol{C}}^\dagger_{\mathcal{Q}^\star\backslash\mathcal{Q}}\tilde{\boldsymbol{C}}_{\bullet\backslash\mathcal{Q}^\star}\right)^t\tilde{\boldsymbol{C}}^t_{\mathcal{Q}^\star\backslash\mathcal{Q}}\boldsymbol{r}_\mathcal{Q}\|_\infty}{\|\tilde{\boldsymbol{C}}^t_{\mathcal{Q}^\star\backslash\mathcal{Q}}\boldsymbol{r}_\mathcal{Q}\|_\infty}.$$

This expression can obviously be majorized using the matrix norm:

$$\phi(\boldsymbol{r}_\mathcal{Q}) \leqslant \|\left(\tilde{\boldsymbol{C}}^\dagger_{\mathcal{Q}^\star\backslash\mathcal{Q}}\tilde{\boldsymbol{C}}_{\bullet\backslash\mathcal{Q}^\star}\right)^t\|_{\infty,\infty}. \tag{15}$$

Since the ℓ_∞ norm of a matrix is equal to the ℓ_1 norm of its transpose and $\|\,.\,\|_{1,1}$ equals the maximum column sum of the absolute value of its argument [1, Theorem 3.1], the upper bound of (15) rereads

$$\max_{j\notin\mathcal{Q}^\star}\|\tilde{\boldsymbol{C}}^\dagger_{\mathcal{Q}^\star\backslash\mathcal{Q}}\tilde{\boldsymbol{c}}_j\|_1 = \max_{j\notin\mathcal{Q}^\star} F^{\text{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$$

according to Lemma 1. By definition of ERC-Oxx($\boldsymbol{A},\mathcal{Q}^\star,\mathcal{Q}$), this upper bound is lower than 1 thus $\phi(\boldsymbol{r}_\mathcal{Q})<1$. ∎

2) Recursive expression of the ERC-Oxx formulas: We elaborate recursive expressions of $F^{\text{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$ when $\mathcal{Q}$ is increased by one element resulting in the new subset $\mathcal{Q}'\subsetneq\mathcal{Q}^\star$ (here, we do not consider the case where $\mathcal{Q}'=\mathcal{Q}^\star$ since $F^{\text{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}^\star}(\boldsymbol{a}_j)$ is not properly defined, (4) and (5) being empty sums). We will use the notation $\mathcal{Q}'=\mathcal{Q}\cup\{\ell\}$ where $\ell\in\mathcal{Q}^\star\backslash\mathcal{Q}$. To avoid any confusion, $\tilde{\boldsymbol{a}}_i$ will be systematically replaced by $\tilde{\boldsymbol{a}}^\mathcal{Q}_i$ and $\tilde{\boldsymbol{a}}^{\mathcal{Q}'}_i$ to express the dependence upon $\mathcal{Q}$ and $\mathcal{Q}'$, respectively. In the same way, $\tilde{\boldsymbol{b}}_i$ will be replaced by $\tilde{\boldsymbol{b}}^\mathcal{Q}_i$ or $\tilde{\boldsymbol{b}}^{\mathcal{Q}'}_i$ but for simplicity, we will keep the matrix notations $\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}$ and $\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star\backslash\mathcal{Q}'}$ without superscript, referring to $\mathcal{Q}$ and $\mathcal{Q}'$, respectively.

Let us first link $\tilde{\boldsymbol{b}}^\mathcal{Q}_i$ to $\tilde{\boldsymbol{b}}^{\mathcal{Q}'}_i$ when $\tilde{\boldsymbol{a}}^{\mathcal{Q}'}_i\neq\boldsymbol{0}$.

Lemma 5 *Assume that $\boldsymbol{A}_{\mathcal{Q}'}$ is full rank and $\mathcal{Q}'=\mathcal{Q}\cup\{\ell\}\subsetneq\mathcal{Q}^\star$. Then, $\operatorname{span}(\boldsymbol{A}_\mathcal{Q})^\perp$ is the orthogonal direct sum of the subspaces $\operatorname{span}(\boldsymbol{A}_{\mathcal{Q}'})^\perp$ and $\operatorname{span}(\tilde{\boldsymbol{a}}^\mathcal{Q}_\ell)$, and the normalized projection of any atom $\boldsymbol{a}_i\notin\operatorname{span}(\boldsymbol{A}_{\mathcal{Q}'})$ takes the form:*

$$\tilde{\boldsymbol{b}}^\mathcal{Q}_i = \eta^{\mathcal{Q},\mathcal{Q}'}_i\tilde{\boldsymbol{b}}^{\mathcal{Q}'}_i + \chi^{\mathcal{Q},\mathcal{Q}'}_i\tilde{\boldsymbol{b}}^\mathcal{Q}_\ell \tag{16}$$

where

$$\eta^{\mathcal{Q},\mathcal{Q}'}_i = \frac{\left\|\tilde{\boldsymbol{a}}^{\mathcal{Q}'}_i\right\|}{\left\|\tilde{\boldsymbol{a}}^\mathcal{Q}_i\right\|} \in (0,1], \tag{17}$$

$$\chi^{\mathcal{Q},\mathcal{Q}'}_i = \langle\tilde{\boldsymbol{b}}^\mathcal{Q}_i,\tilde{\boldsymbol{b}}^\mathcal{Q}_\ell\rangle, \tag{18}$$

$$\left(\eta^{\mathcal{Q},\mathcal{Q}'}_i\right)^2 + \left(\chi^{\mathcal{Q},\mathcal{Q}'}_i\right)^2 = 1. \tag{19}$$

Proof: Since $\mathcal{Q}\subsetneq\mathcal{Q}'$, we have $\operatorname{span}(\boldsymbol{A}_{\mathcal{Q}'})^\perp\subseteq\operatorname{span}(\boldsymbol{A}_\mathcal{Q})^\perp$. Because $\boldsymbol{A}_{\mathcal{Q}'}$ is full rank, $\operatorname{span}(\boldsymbol{A}_{\mathcal{Q}'})^\perp$ and $\operatorname{span}(\boldsymbol{A}_\mathcal{Q})^\perp$ are of consecutive dimensions. Moreover, $\tilde{\boldsymbol{a}}^\mathcal{Q}_\ell = \boldsymbol{a}_\ell - \boldsymbol{P}_\mathcal{Q}\boldsymbol{a}_\ell \in \operatorname{span}(\boldsymbol{A}_{\mathcal{Q}'})\cap\operatorname{span}(\boldsymbol{A}_\mathcal{Q})^\perp$, and $\tilde{\boldsymbol{a}}^\mathcal{Q}_\ell\neq\boldsymbol{0}$ since $\boldsymbol{A}_{\mathcal{Q}'}$ is full rank. As a vector of $\operatorname{span}(\boldsymbol{A}_{\mathcal{Q}'})$, $\tilde{\boldsymbol{a}}^\mathcal{Q}_\ell$ is orthogonal to $\operatorname{span}(\boldsymbol{A}_{\mathcal{Q}'})^\perp$. It follows that $\operatorname{span}(\tilde{\boldsymbol{a}}^\mathcal{Q}_\ell)$ is the orthogonal complement of $\operatorname{span}(\boldsymbol{A}_{\mathcal{Q}'})^\perp$ in $\operatorname{span}(\boldsymbol{A}_\mathcal{Q})^\perp$.

The orthogonal decomposition of $\tilde{\boldsymbol{a}}_i = \boldsymbol{P}_{\mathcal{Q}}^{\perp} \boldsymbol{a}_i$ reads:

$$\tilde{\boldsymbol{a}}_i^{\mathcal{Q}} = \tilde{\boldsymbol{a}}_i^{\mathcal{Q}'} + \langle \tilde{\boldsymbol{a}}_i^{\mathcal{Q}}, \tilde{\boldsymbol{b}}_\ell^{\mathcal{Q}} \rangle \tilde{\boldsymbol{b}}_\ell^{\mathcal{Q}}$$

since $\tilde{\boldsymbol{b}}_\ell^{\mathcal{Q}}$ is of unit norm. Replacing $\tilde{\boldsymbol{a}}_i^{\mathcal{Q}} = \|\tilde{\boldsymbol{a}}_i^{\mathcal{Q}}\| \, \tilde{\boldsymbol{b}}_i^{\mathcal{Q}}$ and $\tilde{\boldsymbol{a}}_i^{\mathcal{Q}'} = \|\tilde{\boldsymbol{a}}_i^{\mathcal{Q}'}\| \, \tilde{\boldsymbol{b}}_i^{\mathcal{Q}'}$ yields (16)-(18). Pythagoras' theorem yields (19). The assumption $\boldsymbol{a}_i \notin \operatorname{span}(\boldsymbol{A}_{\mathcal{Q}'})$ implies that $\tilde{\boldsymbol{a}}_i^{\mathcal{Q}'} \neq \mathbf{0}$, then $\eta_i^{\mathcal{Q},\mathcal{Q}'} > 0$. ■

Lemma 6 *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. Let $\mathcal{Q} \subsetneq \mathcal{Q}' \subsetneq \mathcal{Q}^\star$ with $\mathcal{Q}' = \mathcal{Q} \cup \{\ell\}$. Then, $\operatorname{span}(\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star \setminus \mathcal{Q}})$ is the orthogonal direct sum of $\operatorname{span}(\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star \setminus \mathcal{Q}'})$ and $\operatorname{span}(\tilde{\boldsymbol{b}}_\ell^{\mathcal{Q}})$.*

Proof: According to Corollary 3 in Appendix B, $\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star \setminus \mathcal{Q}}$ and $\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star \setminus \mathcal{Q}'}$ are full rank matrices, thus their column spans are of consecutive cardinalities. Lemma 5 states that $\tilde{\boldsymbol{b}}_\ell^{\mathcal{Q}}$ is orthogonal to $\operatorname{span}(\boldsymbol{A}_{\mathcal{Q}'})^\perp$, thus it is orthogonal to $\tilde{\boldsymbol{b}}_i^{\mathcal{Q}'} \in \operatorname{span}(\boldsymbol{A}_{\mathcal{Q}'})^\perp$ for all $i \in \mathcal{Q}^\star \setminus \mathcal{Q}'$. ■

We finally establish a link between $F^{\text{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$ and $F^{\text{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}'}(\boldsymbol{a}_j)$. It is a simple recursive relation in the case of OMP. For OLS, we cannot directly relate the two quantities but we express $F^{\text{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) = \big\|\tilde{\boldsymbol{B}}^\dagger_{\mathcal{Q}^\star \setminus \mathcal{Q}} \tilde{\boldsymbol{b}}_j^{\mathcal{Q}}\big\|_1$ with respect to $\tilde{\boldsymbol{B}}^\dagger_{\mathcal{Q}^\star \setminus \mathcal{Q}'} \tilde{\boldsymbol{b}}_j^{\mathcal{Q}'}$.

Lemma 7 *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. Let $\mathcal{Q} \subsetneq \mathcal{Q}' \subsetneq \mathcal{Q}^\star$ with $\mathcal{Q}' = \mathcal{Q} \cup \{\ell\}$ and let $j \notin \mathcal{Q}^\star$. If $\boldsymbol{a}_j \notin \operatorname{span}(\boldsymbol{A}_{\mathcal{Q}'})$, then $F^{\text{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}'}(\boldsymbol{a}_j)$ takes the forms* (20) *and* (21) *where $\eta_i^{\mathcal{Q},\mathcal{Q}'}$ and $\chi_i^{\mathcal{Q},\mathcal{Q}'}$ are defined in* (17)-(18) *and $\boldsymbol{\beta}_j^{\mathcal{Q}^\star \setminus \mathcal{Q}'} \triangleq \tilde{\boldsymbol{B}}^\dagger_{\mathcal{Q}^\star \setminus \mathcal{Q}'} \tilde{\boldsymbol{b}}_j^{\mathcal{Q}'}$.*

Proof: (20) straightforwardly follows from the definition (4) of $F^{\text{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$.

Let us now establish (21). We denote by $\tilde{\boldsymbol{P}}_{\mathcal{Q}^\star \setminus \mathcal{Q}}$ and $\tilde{\boldsymbol{P}}_{\mathcal{Q}^\star \setminus \mathcal{Q}'}$ the orthogonal projectors on $\operatorname{span}(\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star \setminus \mathcal{Q}})$ and $\operatorname{span}(\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star \setminus \mathcal{Q}'})$. Because $\operatorname{span}(\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star \setminus \mathcal{Q}})$ is the orthogonal direct sum of $\operatorname{span}(\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star \setminus \mathcal{Q}'})$ and $\operatorname{span}(\tilde{\boldsymbol{b}}_\ell^{\mathcal{Q}})$ (Lemma 6), we have the orthogonal decomposition:

$$\tilde{\boldsymbol{P}}_{\mathcal{Q}^\star \setminus \mathcal{Q}} \tilde{\boldsymbol{b}}_j^{\mathcal{Q}} = \tilde{\boldsymbol{P}}_{\mathcal{Q}^\star \setminus \mathcal{Q}'} \tilde{\boldsymbol{b}}_j^{\mathcal{Q}} + \chi_j^{\mathcal{Q},\mathcal{Q}'} \tilde{\boldsymbol{b}}_\ell^{\mathcal{Q}}.$$

(16) yields

$$\tilde{\boldsymbol{P}}_{\mathcal{Q}^\star \setminus \mathcal{Q}} \tilde{\boldsymbol{b}}_j^{\mathcal{Q}} = \eta_j^{\mathcal{Q},\mathcal{Q}'} \tilde{\boldsymbol{P}}_{\mathcal{Q}^\star \setminus \mathcal{Q}'} \tilde{\boldsymbol{b}}_j^{\mathcal{Q}'} + \chi_j^{\mathcal{Q},\mathcal{Q}'} \tilde{\boldsymbol{b}}_\ell^{\mathcal{Q}}$$

($\tilde{\boldsymbol{P}}_{\mathcal{Q}^\star \setminus \mathcal{Q}'} \tilde{\boldsymbol{b}}_\ell^{\mathcal{Q}} = \mathbf{0}$ according to Lemma 6) and then

$$\tilde{\boldsymbol{P}}_{\mathcal{Q}^\star \setminus \mathcal{Q}} \tilde{\boldsymbol{b}}_j^{\mathcal{Q}} = \eta_j^{\mathcal{Q},\mathcal{Q}'} \sum_{i \in \mathcal{Q}^\star \setminus \mathcal{Q}'} \boldsymbol{\beta}_j^{\mathcal{Q}^\star \setminus \mathcal{Q}'}(i) \tilde{\boldsymbol{b}}_i^{\mathcal{Q}'} + \chi_j^{\mathcal{Q},\mathcal{Q}'} \tilde{\boldsymbol{b}}_\ell^{\mathcal{Q}}$$

by definition of $\boldsymbol{\beta}_j^{\mathcal{Q}^\star \setminus \mathcal{Q}'}$. In the latter equation, we re-express $\tilde{\boldsymbol{b}}_i^{\mathcal{Q}'}$ with respect to $\tilde{\boldsymbol{b}}_i^{\mathcal{Q}}$ using (16):

$$\tilde{\boldsymbol{P}}_{\mathcal{Q}^\star \setminus \mathcal{Q}} \tilde{\boldsymbol{b}}_j^{\mathcal{Q}} = \eta_j^{\mathcal{Q},\mathcal{Q}'} \sum_{i \in \mathcal{Q}^\star \setminus \mathcal{Q}'} \frac{\boldsymbol{\beta}_j^{\mathcal{Q}^\star \setminus \mathcal{Q}'}(i)}{\eta_i^{\mathcal{Q},\mathcal{Q}'}} \tilde{\boldsymbol{b}}_i^{\mathcal{Q}} + \left\{ \chi_j^{\mathcal{Q},\mathcal{Q}'} - \eta_j^{\mathcal{Q},\mathcal{Q}'} \sum_{i \in \mathcal{Q}^\star \setminus \mathcal{Q}'} \frac{\boldsymbol{\beta}_j^{\mathcal{Q}^\star \setminus \mathcal{Q}'}(i) \chi_i^{\mathcal{Q},\mathcal{Q}'}}{\eta_i^{\mathcal{Q},\mathcal{Q}'}} \right\} \tilde{\boldsymbol{b}}_\ell^{\mathcal{Q}}.$$

Thus, $F^{\text{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) = \big\|\tilde{\boldsymbol{B}}^\dagger_{\mathcal{Q}^\star \setminus \mathcal{Q}} \tilde{\boldsymbol{b}}_j^{\mathcal{Q}}\big\|_1$ reads (21). ■

3) The ERC is a sufficient recovery condition for OLS: The key result of Lemma 2 (see Section III-D) states that when $j \notin \mathcal{Q}^\star$, $F^{\text{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$ is decreasing when $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ is growing provided that $F^{\text{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) < 1$, and that $F^{\text{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$ is always decreasing.

Proof of Lemma 2: It is sufficient to prove the result when $\operatorname{Card}[\mathcal{Q}'] = \operatorname{Card}[\mathcal{Q}] + 1$. The case $\operatorname{Card}[\mathcal{Q}'] > \operatorname{Card}[\mathcal{Q}] + 1$ obviously deduces from the former case by recursion.

Let $\mathcal{Q} \subsetneq \mathcal{Q}' \subsetneq \mathcal{Q}^\star$ with $\operatorname{Card}[\mathcal{Q}'] = \operatorname{Card}[\mathcal{Q}]+1$. The result is obvious when $\boldsymbol{a}_j \in \operatorname{span}(\boldsymbol{A}_{\mathcal{Q}'})$: $\tilde{\boldsymbol{a}}_j = \mathbf{0}$ then $F^{\text{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}'}(\boldsymbol{a}_j) = 0$. When $\boldsymbol{a}_j \notin \operatorname{span}(\boldsymbol{A}_{\mathcal{Q}'})$, (8) obviously deduces from (20). The proof of (9) relies on the study of function $\varphi(\eta) = |\sqrt{1-\eta^2} - C\eta| + D\eta$ which is fully defined in (27), (28) and (29) in Appendix C. Because this study is rather technical, we place it in Appendix C.

We notice that $F^{\text{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$ given in (21) takes the form $\varphi\big(\eta_j^{\mathcal{Q},\mathcal{Q}'}\big)$ where the variables occurring in C and D (see (28) and (29)) are set to $N \leftarrow \operatorname{Card}[\mathcal{Q}^\star \setminus \mathcal{Q}']$, $\eta_i \leftarrow \eta_i^{\mathcal{Q},\mathcal{Q}'}$, $\chi_i \leftarrow \chi_i^{\mathcal{Q},\mathcal{Q}'}$, and $\boldsymbol{\beta} \leftarrow \operatorname{sgn}\big(\chi_j^{\mathcal{Q},\mathcal{Q}'}\big) \boldsymbol{\beta}_j^{\mathcal{Q}^\star \setminus \mathcal{Q}'}$. Now, we invoke Lemma 14 in Appendix C: as $F^{\text{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}'}(\boldsymbol{a}_j) = \big\|\boldsymbol{\beta}_j^{\mathcal{Q}^\star \setminus \mathcal{Q}'}\big\|_1$ plays the role of $\|\boldsymbol{\beta}\|_1$, $F^{\text{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) < 1$ implies that $F^{\text{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}'}(\boldsymbol{a}_j) \leqslant F^{\text{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$. ■

We deduce from Lemmas 2 and 4 that ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) are sufficient recovery conditions when $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ has been reached (Theorem 3).

Proof of Theorem 3: Apply Lemma 4 at each iteration $q, \ldots, k-1$ until the increased subset $\mathcal{Q}'$ matches $\mathcal{Q}^\star$. The ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, .$) assumption of Lemma 4 is always fulfilled according to Lemma 2. ■

Finally, we prove that ERC($\boldsymbol{A}, \mathcal{Q}^\star$) is a necessary and sufficient condition of successful recovery for OLS (Theorem 2).

Proof of Theorem 2: The sufficient condition is a special case of Theorem 3 for $\mathcal{Q} = \emptyset$. The necessary condition identifies with that of Theorem 1 since ERC-OLS($\boldsymbol{A}, \mathcal{Q}^\star, \emptyset$) simplifies to ERC($\boldsymbol{A}, \mathcal{Q}^\star$). ■

B. Necessary conditions

We provide the technical analysis to prove that ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) is not only a sufficient condition of exact recovery when $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ has been reached, but also a necessary condition in the worst case. We will prove Theorems 4 and 5 (see Section III) generalizing Tropp's necessary condition [1, Theorem 3.10] to any iteration of OMP and OLS.

We will first assume that Oxx exactly recovers $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ in $q = \mathrm{Card}\,[\mathcal{Q}]$ iterations with some input vector in $\mathrm{span}(\boldsymbol{A}_\mathcal{Q})$. This reachability assumption allows us to carry out a parallel analysis of OMP and OLS (subsection A-B1) leading to the following proposition.

Proposition 1 ***[Necessary condition for Oxx after*** q ***iterations]*** *Assume that* $\boldsymbol{A}_{\mathcal{Q}^\star}$ *is full rank and* $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ *is reachable from an input in* $\mathrm{span}(\boldsymbol{A}_\mathcal{Q})$ *by Oxx. If ERC-Oxx(*$\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$*) does not hold, then there exists* $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ *for which Oxx selects* $\mathcal{Q}$ *in the first* q *iterations and then a wrong atom at iteration* $(q+1)$.

This proposition coincides with Theorem 5 in the case of OMP whereas for OLS, Theorem 4 does not require the assumption that $\mathcal{Q}$ is reachable (subsection A-B2).

1) Parallel analysis of OMP and OLS: *Proof of Proposition 1:* We proceed the proof of Lemma 4 backwards. By assumption, the right hand-side of inequality (15) is equal to

$$\|(\tilde{\boldsymbol{C}}^\dagger_{\mathcal{Q}^\star\backslash\mathcal{Q}}\tilde{\boldsymbol{C}}_{\bullet\backslash\mathcal{Q}^\star})^t\|_{\infty,\infty} = \max_{j\notin\mathcal{Q}^\star} F^{\mathrm{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j) \geqslant 1.$$

By definition of induced norms, there exists a vector $\boldsymbol{v} \in \mathbb{R}^{k-q}$ satisfying $\boldsymbol{v} \neq \boldsymbol{0}$ and

$$\frac{\|(\tilde{\boldsymbol{C}}^\dagger_{\mathcal{Q}^\star\backslash\mathcal{Q}}\tilde{\boldsymbol{C}}_{\bullet\backslash\mathcal{Q}^\star})^t\boldsymbol{v}\|_\infty}{\|\boldsymbol{v}\|_\infty} = \|(\tilde{\boldsymbol{C}}^\dagger_{\mathcal{Q}^\star\backslash\mathcal{Q}}\tilde{\boldsymbol{C}}_{\bullet\backslash\mathcal{Q}^\star})^t\|_{\infty,\infty} \geqslant 1. \tag{22}$$

Define

$$\hat{\boldsymbol{y}} = \boldsymbol{A}_{\mathcal{Q}^\star\backslash\mathcal{Q}}(\tilde{\boldsymbol{C}}^t_{\mathcal{Q}^\star\backslash\mathcal{Q}}\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star\backslash\mathcal{Q}})^{-1}\boldsymbol{v}. \tag{23}$$

The matrix inversion in (23) is well defined since $\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}$ is full rank (Corollary 3 in Appendix B) and $\tilde{\boldsymbol{C}}_{\mathcal{Q}^\star\backslash\mathcal{Q}} = \tilde{\boldsymbol{A}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}$ or $\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}$ reads as the right product of $\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}$ with a nondegenerate diagonal matrix. By taking into account that $\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star\backslash\mathcal{Q}} = \boldsymbol{P}^\perp_\mathcal{Q}\boldsymbol{A}_{\mathcal{Q}^\star\backslash\mathcal{Q}}$, we obtain that

$$\boldsymbol{v} = \tilde{\boldsymbol{C}}^t_{\mathcal{Q}^\star\backslash\mathcal{Q}}\boldsymbol{P}^\perp_\mathcal{Q}\hat{\boldsymbol{y}}. \tag{24}$$

Since the left hand-side of (22) identifies with $\phi(\boldsymbol{P}^\perp_\mathcal{Q}\hat{\boldsymbol{y}})$ where ϕ is defined in (14), (22) yields:

$$\max_{j\notin\mathcal{Q}^\star}|\langle\boldsymbol{P}^\perp_\mathcal{Q}\hat{\boldsymbol{y}},\tilde{\boldsymbol{c}}_j\rangle| \geqslant \max_{i\in\mathcal{Q}^\star\backslash\mathcal{Q}}|\langle\boldsymbol{P}^\perp_\mathcal{Q}\hat{\boldsymbol{y}},\tilde{\boldsymbol{c}}_i\rangle|. \tag{25}$$

Moreover, we have $\boldsymbol{P}^\perp_\mathcal{Q}\hat{\boldsymbol{y}} \neq \boldsymbol{0}$ according to (24) and $\boldsymbol{v} \neq \boldsymbol{0}$.

Now, let $\boldsymbol{z} \in \mathrm{span}(\boldsymbol{A}_\mathcal{Q})$ denote the input for which Oxx recovers $\mathcal{Q}$. According to Lemma 15 in Appendix D, the first q iterations of Oxx with the modified input $\boldsymbol{y} = \boldsymbol{z}+\varepsilon\hat{\boldsymbol{y}}$ also select $\mathcal{Q}$ when $\varepsilon > 0$ is sufficiently small. Because $\boldsymbol{P}^\perp_\mathcal{Q}\boldsymbol{y} = \varepsilon\boldsymbol{P}^\perp_\mathcal{Q}\hat{\boldsymbol{y}}$ and (25) holds, the $(q+1)$-th iteration of Oxx necessarily selects a wrong atom. ■

At this point, we have proved Theorem 5 which is relative to OMP.

2) OLS ability to reach any subset: In order to prove Theorem 4, we establish that any subset $\mathcal{Q}$ can be reached using OLS with some input $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_\mathcal{Q})$ (Lemma 3). To generate $\boldsymbol{y}$, we assign decreasing weight coefficients to the atoms $\{\boldsymbol{a}_i,\, i \in \mathcal{Q}\}$ with a rate of decrease which is high enough.

Proof of Lemma 3: Without loss of generality, we assume that the elements of $\mathcal{Q}$ correspond to the first q atoms. For arbitrary values of $\varepsilon_2,\dots,\varepsilon_q > 0$, we define the following recursive construction:

- $\boldsymbol{y}_1 = \boldsymbol{a}_1$,
- $\boldsymbol{y}_p = \boldsymbol{y}_{p-1} + \varepsilon_p\boldsymbol{a}_p$ for $p \in \{2,\dots,q\}$.

($\boldsymbol{y}_p$ implicitly depends on $\varepsilon_2,\dots,\varepsilon_p$) and set $\boldsymbol{y} \triangleq \boldsymbol{y}_q$. We show by recursion that there exist $\varepsilon_2,\dots,\varepsilon_p > 0$ such that OLS with $\boldsymbol{y}_p$ as input successively selects $\boldsymbol{a}_1,\dots,\boldsymbol{a}_p$ during the first p iterations (in particular, the selection rule (2) always yields a unique maximum).

The statement is obviously true for $\boldsymbol{y}_1 = \boldsymbol{a}_1$. Assume that it is true for $\boldsymbol{y}_{p-1}$ with some $\varepsilon_2,\dots,\varepsilon_{p-1} > 0$ (these parameters will remain fixed in the following). According to Lemma 15 in Appendix D, there exists $\varepsilon_p > 0$ such that OLS with $\boldsymbol{y}_p = \boldsymbol{y}_{p-1} + \varepsilon_p\boldsymbol{a}_p$ as input selects the same atoms as with $\boldsymbol{y}_{p-1}$ during the first $p-1$ iterations, *i.e.*, $\boldsymbol{a}_1,\dots,\boldsymbol{a}_{p-1}$ are successively chosen. At iteration p, the current active set thus reads $\mathcal{Q}' = \{1,\dots,p-1\}$ and the OLS residual corresponding to $\boldsymbol{y}_p$ takes the form

$$\boldsymbol{r}_{\mathcal{Q}'} = \boldsymbol{P}^\perp_{\mathcal{Q}'}\boldsymbol{y}_{p-1} + \varepsilon_i\boldsymbol{P}^\perp_{\mathcal{Q}'}\boldsymbol{a}_p \;=\; \varepsilon_p\tilde{\boldsymbol{a}}^{\mathcal{Q}'}_p$$

since $\boldsymbol{y}_{p-1} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}'})$. Thus, $\boldsymbol{r}_{\mathcal{Q}'}$ is proportional to $\tilde{\boldsymbol{a}}^{\mathcal{Q}'}_p$ and then to $\tilde{\boldsymbol{b}}^{\mathcal{Q}'}_p$. Finally, the OLS criterion (2) is maximum for the atom $\boldsymbol{a}_p$ and the maximum value is equal to $|\langle\boldsymbol{r}_{\mathcal{Q}'},\tilde{\boldsymbol{b}}^{\mathcal{Q}'}_p\rangle| = \|\boldsymbol{r}_{\mathcal{Q}'}\|$ since $\tilde{\boldsymbol{b}}^{\mathcal{Q}'}_p$ is of unit norm.

Finally, we show that no other atom $\boldsymbol{a}_i$ yields this maximum value. Apply Lemma 8 in Appendix B: the full rankness of $\boldsymbol{A}_{\mathcal{Q}'\cup\{p,i\}}$ (as a family of less than $\mathrm{spark}(\boldsymbol{A})$ atoms) implies that $[\tilde{\boldsymbol{b}}^{\mathcal{Q}'}_p,\tilde{\boldsymbol{b}}^{\mathcal{Q}'}_i]$ is full rank, thus $\tilde{\boldsymbol{b}}^{\mathcal{Q}'}_p$ and $\tilde{\boldsymbol{b}}^{\mathcal{Q}'}_i$ cannot be collinear. ■

Using Lemma 3, Proposition 1 simplifies to Theorem 4 in which the assumption that $\mathcal{Q}$ is reachable by OLS is omitted.

APPENDIX B
RE-EXPRESSION OF THE ERC-OXX FORMULAS

In this appendix, we prove Lemma 1 by successively re-expressing $\tilde{\boldsymbol{A}}^\dagger_{\mathcal{Q}^\star\backslash\mathcal{Q}}\tilde{\boldsymbol{a}}_j$ and $\tilde{\boldsymbol{B}}^\dagger_{\mathcal{Q}^\star\backslash\mathcal{Q}}\tilde{\boldsymbol{b}}_j$. Let us first show that when $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank, the matrices $\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}$ and $\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}$ are full rank. This result is stated below as a corollary of Lemma 8.

Lemma 8 *If $\mathcal{Q}\cap\mathcal{Q}'=\emptyset$ and $\boldsymbol{A}_{\mathcal{Q}\cup\mathcal{Q}'}$ is full rank, then $\tilde{\boldsymbol{A}}^{\mathcal{Q}}_{\mathcal{Q}'}$ and $\tilde{\boldsymbol{B}}^{\mathcal{Q}}_{\mathcal{Q}'}$ are full rank.*

Proof: To prove that $\tilde{\boldsymbol{A}}^{\mathcal{Q}}_{\mathcal{Q}'}$ is full rank, we assume that $\sum_{i\in\mathcal{Q}'}\alpha_i\tilde{\boldsymbol{a}}^{\mathcal{Q}}_i=\boldsymbol{0}$ with $\alpha_i\in\mathbb{R}$. By definition of $\tilde{\boldsymbol{a}}^{\mathcal{Q}}_i=\boldsymbol{P}^\perp_{\mathcal{Q}}\boldsymbol{a}_i=\boldsymbol{a}_i-\boldsymbol{P}_{\mathcal{Q}}\boldsymbol{a}_i$, it follows that $\sum_{i\in\mathcal{Q}'}\alpha_i\boldsymbol{a}_i\in\operatorname{span}(\boldsymbol{A}_{\mathcal{Q}})$. Since $\boldsymbol{A}_{\mathcal{Q}\cup\mathcal{Q}'}$ is full rank, we conclude that all α_i's are 0.

The full rankness of $\tilde{\boldsymbol{B}}^{\mathcal{Q}}_{\mathcal{Q}'}$ follows from that of $\tilde{\boldsymbol{A}}^{\mathcal{Q}}_{\mathcal{Q}'}$ since for all $i\in\mathcal{Q}'$, $\tilde{\boldsymbol{b}}^{\mathcal{Q}}_i=\tilde{\boldsymbol{a}}^{\mathcal{Q}}_i/\|\tilde{\boldsymbol{a}}^{\mathcal{Q}}_i\|$ is collinear to $\tilde{\boldsymbol{a}}^{\mathcal{Q}}_i$. ■

The application of Lemma 8 to $\mathcal{Q}'=\mathcal{Q}^\star\backslash\mathcal{Q}$ leads to the following corollary.

Corollary 3 *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. For $\mathcal{Q}\subsetneq\mathcal{Q}^\star$, $\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}$ and $\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}$ are full rank.*

Lemma 9 *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. For $\mathcal{Q}\subsetneq\mathcal{Q}^\star$ and $j\notin\mathcal{Q}^\star$, $\tilde{\boldsymbol{A}}^\dagger_{\mathcal{Q}^\star\backslash\mathcal{Q}}\tilde{\boldsymbol{a}}_j=\big(\boldsymbol{A}^\dagger_{\mathcal{Q}^\star}\boldsymbol{a}_j\big)_{|(\mathcal{Q}^\star\backslash\mathcal{Q})}$ where $|$ denotes the restriction of a vector to a subset of its coefficients.*

Proof: The orthogonal decomposition of $\boldsymbol{a}_j$ on $\operatorname{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ takes the form:

$$\boldsymbol{a}_j=\boldsymbol{A}_{\mathcal{Q}^\star}\big(\boldsymbol{A}^\dagger_{\mathcal{Q}^\star}\boldsymbol{a}_j\big)+\boldsymbol{P}^\perp_{\mathcal{Q}^\star}\boldsymbol{a}_j.$$

Projecting onto $\operatorname{span}(\boldsymbol{A}_{\mathcal{Q}})^\perp$, we obtain

$$\tilde{\boldsymbol{a}}_j=\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}\big(\boldsymbol{A}^\dagger_{\mathcal{Q}^\star}\boldsymbol{a}_j\big)_{|(\mathcal{Q}^\star\backslash\mathcal{Q})}+\boldsymbol{P}^\perp_{\mathcal{Q}^\star}\boldsymbol{a}_j \tag{26}$$

($\boldsymbol{P}^\perp_{\mathcal{Q}}\boldsymbol{P}^\perp_{\mathcal{Q}^\star}=\boldsymbol{P}^\perp_{\mathcal{Q}^\star}$ because $\operatorname{span}(\boldsymbol{A}_{\mathcal{Q}^\star})^\perp\subseteq\operatorname{span}(\boldsymbol{A}_{\mathcal{Q}})^\perp$). For $i\in\mathcal{Q}^\star\backslash\mathcal{Q}$, $\tilde{\boldsymbol{a}}_i=\boldsymbol{a}_i-\boldsymbol{P}_{\mathcal{Q}}\boldsymbol{a}_i\in\operatorname{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$. Thus, we have $\operatorname{span}(\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star\backslash\mathcal{Q}})\subseteq\operatorname{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$, and $\boldsymbol{P}^\perp_{\mathcal{Q}^\star}\boldsymbol{a}_j$ is orthogonal to $\operatorname{span}(\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star\backslash\mathcal{Q}})$. According to Corollary 3, $\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}$ is full rank. It follows from (26) that $\tilde{\boldsymbol{A}}^\dagger_{\mathcal{Q}^\star\backslash\mathcal{Q}}\tilde{\boldsymbol{a}}_j=\big(\boldsymbol{A}^\dagger_{\mathcal{Q}^\star}\boldsymbol{a}_j\big)_{|(\mathcal{Q}^\star\backslash\mathcal{Q})}$. ■

Lemma 10 *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. For $\mathcal{Q}\subsetneq\mathcal{Q}^\star$ and $j\notin\mathcal{Q}^\star$,*

$$\|\tilde{\boldsymbol{a}}_j\|\,\tilde{\boldsymbol{B}}^\dagger_{\mathcal{Q}^\star\backslash\mathcal{Q}}\tilde{\boldsymbol{b}}_j=\boldsymbol{\Delta}_{\|\tilde{\boldsymbol{a}}_i\|}\big(\boldsymbol{A}^\dagger_{\mathcal{Q}^\star}\boldsymbol{a}_j\big)_{|(\mathcal{Q}^\star\backslash\mathcal{Q})}$$

where $\boldsymbol{\Delta}_{\|\tilde{\boldsymbol{a}}_i\|}$ stands for the diagonal matrix whose diagonal elements are $\{\|\tilde{\boldsymbol{a}}_i\|,\,i\in\mathcal{Q}^\star\backslash\mathcal{Q}\}$.

Proof: The result directly follows from $\tilde{\boldsymbol{a}}_j=\|\tilde{\boldsymbol{a}}_j\|\,\tilde{\boldsymbol{b}}_j$, $\tilde{\boldsymbol{b}}_i=\tilde{\boldsymbol{a}}_i/\|\tilde{\boldsymbol{a}}_i\|$ for $i\in\mathcal{Q}^\star\backslash\mathcal{Q}$, and from Lemma 9. ■

Proof of Lemma 1: The result is obvious when $\tilde{\boldsymbol{a}}_j=\boldsymbol{0}$. It follows from Lemmas 9 and 10 when $\tilde{\boldsymbol{a}}_j\neq\boldsymbol{0}$. ■

APPENDIX C
TECHNICAL RESULTS NEEDED FOR THE PROOF OF LEMMA 2

With simplified notations, the expression (21) of $F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_j)$ reads

$$\varphi(\eta)\triangleq|\sqrt{1-\eta^2}-C\eta|+D\eta \tag{27}$$

where $\eta\in(0,1]$ and C and D take the form

$$C=\sum_{i=1}^{N}\frac{\beta_i\chi_i}{\eta_i} \tag{28}$$

$$D=\sum_{i=1}^{N}\frac{|\beta_i|}{\eta_i} \tag{29}$$

with $N\geqslant 1$, $\boldsymbol{\beta}=[\beta_1,\ldots,\beta_N]\in\mathbb{R}^N$, and for all i, $\eta_i\in(0,1]$ and $\chi_i\in[-1,1]$ satisfy $\eta_i^2+\chi_i^2=1$. Note that we can freely assume from (21) that $\chi_j^{\mathcal{Q},\mathcal{Q}'}=\pm\sqrt{1-\big(\eta_j^{\mathcal{Q},\mathcal{Q}'}\big)^2}\geqslant 0$. When $\chi_j^{\mathcal{Q},\mathcal{Q}'}<0$, one just needs to replace $\boldsymbol{\beta}$ by $-\boldsymbol{\beta}$ in (28) and (29).

The succession of small lemmas hereafter aims at minorizing $\varphi(\eta)$ for arbitrary values of η, η_i, χ_i and $\boldsymbol{\beta}$. They lead to the main minoration result of Lemma 14.

Lemma 11 *Let $\boldsymbol{\beta}\in\mathbb{R}^N$.*

$$\text{If } C\leqslant 0, \forall\eta\in[0,1],\ \varphi(\eta)\geqslant 1+(\|\boldsymbol{\beta}\|_1-1)\eta. \tag{30}$$

$$\text{If } C>0,\ \min_{\eta\in[0,1]}\varphi(\eta)=\min\Big(1,D/\sqrt{1+C^2}\Big). \tag{31}$$

Proof: We first study the function $f(\eta)\triangleq\sqrt{1-\eta^2}-C\eta$. We have $f(0)=1$, $f(1)=-C$, and f is concave on $[0,1]$. To minorize $\varphi(\eta)=|f(\eta)|+D\eta$, we distinguish two cases depending on the sign of C.

When $C\leqslant 0$, $f(\eta)\geqslant 0$ for all η. Since $|f|=f$ is concave, it can be minorized by the secant line joining $f(0)$ and $f(1)$, therefore, $|f(\eta)|\geqslant 1-(C+1)\eta\geqslant 1-\eta$. (30) follows from $\varphi(\eta)=|f(\eta)|+D\eta$ and $D\geqslant\|\boldsymbol{\beta}\|_1$ (because η_i are all in $(0,1]$).

When $C>0$, $f(\eta)\geqslant 0$ for $\eta\in[0,z]$ and <0 in $(z,1]$, with $z\triangleq 1/\sqrt{1+C^2}$. $D\geqslant 0$ and $f(z)=0$ imply that for $\eta>z$, $\varphi(\eta)\geqslant\varphi(z)$, thus the minimum of φ is reached for $\eta\in[0,z]$. On $[0,z]$, $\varphi(\eta)=f(\eta)+D\eta$ is concave, therefore the minimum value is either $\varphi(0)=1$ or $\varphi(z)=Dz$. ■

The following two lemmas are simple inequalities linking C, D, and $\|\boldsymbol{\beta}\|_1$.

Lemma 12 $\forall \boldsymbol{\beta} \in \mathbb{R}^N$, $D^2 - C^2 \geqslant \|\boldsymbol{\beta}\|_1^2$.

Proof: By developing C^2 and D^2 from (28) and (29), we get

$$C^2 = \sum_i \frac{\beta_i^2 \chi_i^2}{\eta_i^2} + \sum_{i \neq j} \frac{\beta_i \beta_j \chi_i \chi_j}{\eta_i \eta_j}$$

$$D^2 = \sum_i \frac{\beta_i^2}{\eta_i^2} + \sum_{i \neq j} \frac{|\beta_i \beta_j|}{\eta_i \eta_j}$$

Since $\forall i$, $\eta_i^2 + \chi_i^2 = 1$, we have:

$$D^2 - C^2 = \sum_i \beta_i^2 + \sum_{i \neq j} \frac{|\beta_i \beta_j|}{\eta_i \eta_j} (1 - \sigma_i \sigma_j \chi_i \chi_j)$$

$$= \left[\sum_i |\beta_i| \right]^2 + \sum_{i \neq j} |\beta_i \beta_j| \left[\frac{1 - \sigma_i \sigma_j \chi_i \chi_j}{\eta_i \eta_j} - 1 \right] \quad (32)$$

with $\sigma_i = \operatorname{sgn}(\beta_i) = \pm 1$ if $\beta_i \neq 0$, and $\sigma_i = 1$ otherwise. Because η_i and χ_i satisfy $\eta_i^2 + \chi_i^2 = 1$, they reread $\eta_i = \cos \theta_i$ and $\chi_i = \sin \theta_i$, so $\eta_i \eta_j + \sigma_i \sigma_j \chi_i \chi_j = \cos(\theta_i \pm \theta_j) \leqslant 1$ which proves that the last bracketed expression in (32) is non-negative. (32) yields $D^2 - C^2 \geqslant \|\boldsymbol{\beta}\|_1^2$. ■

Lemma 13 $\forall \boldsymbol{\beta} \in \mathbb{R}^N$, $\|\boldsymbol{\beta}\|_1 \leqslant 1$ *implies that* $\|\boldsymbol{\beta}\|_1 \leqslant D/\sqrt{1 + C^2}$.

Proof: $(1 + C^2)\|\boldsymbol{\beta}\|_1^2 \leqslant \|\boldsymbol{\beta}\|_1^2 + C^2 \leqslant D^2$ according to Lemma 12. ■

We can now establish the main lemma that will enable us to conclude that if $F^{\mathrm{OLS}}_{\mathcal{Q}^\star, \mathcal{Q}}(\boldsymbol{a}_j) < 1$, $F^{\mathrm{OLS}}_{\mathcal{Q}^\star, \mathcal{Q}'}(\boldsymbol{a}_j)$ is monotonically nonincreasing when $\mathcal{Q}' \supsetneq \mathcal{Q}$ is growing.

Lemma 14 $\forall \boldsymbol{\beta} \in \mathbb{R}^N$, $\forall \eta \in [0, 1]$, $\varphi(\eta) < 1$ *implies that* $\|\boldsymbol{\beta}\|_1 \leqslant \varphi(\eta)$.

Proof: Apply Lemma 11.

When $C \leqslant 0$, (30) and $\varphi(\eta) < 1$ imply that $(\|\boldsymbol{\beta}\|_1 - 1) < 0$. Since $\eta \leqslant 1$, the lower bound of (30) is larger than $1 + (\|\boldsymbol{\beta}\|_1 - 1) = \|\boldsymbol{\beta}\|_1$.

When $C > 0$, (31) and $\varphi(\eta) < 1$ imply that the minimum value of φ on $[0, 1]$ is $D/\sqrt{1 + C^2} < 1$, then $D^2 - C^2 < 1$. Lemmas 12 and 13 imply that $\|\boldsymbol{\beta}\|_1 \leqslant 1$ and then $\|\boldsymbol{\beta}\|_1 \leqslant D/\sqrt{1 + C^2} \leqslant \varphi(\eta)$. ■

APPENDIX D
BEHAVIOR OF OXX WHEN THE INPUT VECTOR IS SLIGHTLY MODIFIED

Lemma 15 *Let* $\boldsymbol{y}_1$ *and* $\boldsymbol{y}_2 \in \mathbb{R}^m$. *Assume that the selection rule* (1)-(2) *of Oxx with* $\boldsymbol{y}_1$ *as input is strict in the first* $q > 0$ *iterations (the maximizer is unique). Then, when* $\varepsilon > 0$ *is sufficiently small, Oxx selects the same atoms with* $\boldsymbol{y}(\varepsilon) = \boldsymbol{y}_1 + \varepsilon \boldsymbol{y}_2$ *as with* $\boldsymbol{y}_1$ *in the first* q *iterations.*

Proof: We show by recursion that there exists $\varepsilon_p > 0$ such that the first p iterations of Oxx ($p = 1, \ldots, q$) with $\boldsymbol{y}(\varepsilon)$ and $\boldsymbol{y}_1$ as inputs yield the same atoms whenever $\varepsilon < \varepsilon_p$.

Let $p \geqslant 1$. We denote by $\mathcal{Q}$ the subset of cardinality $p - 1$ delivered by Oxx with $\boldsymbol{y}_1$ as input after $p - 1$ iterations. By assumption, $\mathcal{Q}$ is also yielded with $\boldsymbol{y}(\varepsilon)$ when $\varepsilon < \varepsilon_{p-1}$. Since $\boldsymbol{y}(\varepsilon) = \boldsymbol{y}_1 + \varepsilon \boldsymbol{y}_2$, the Oxx residual takes the form $\boldsymbol{r}_{\mathcal{Q}} = \boldsymbol{r}_1 + \varepsilon \boldsymbol{r}_2$ where $\boldsymbol{r}_{\mathcal{Q}}$, $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are obtained by projecting $\boldsymbol{y}(\varepsilon)$, $\boldsymbol{y}_1$, and $\boldsymbol{y}_2$, respectively onto $\operatorname{span}(\boldsymbol{A}_{\mathcal{Q}})^\perp$. Hence, for $i \notin \mathcal{Q}$,

$$\langle \boldsymbol{r}_{\mathcal{Q}}, \tilde{\boldsymbol{c}}_i \rangle = \langle \boldsymbol{r}_1, \tilde{\boldsymbol{c}}_i \rangle + \varepsilon \langle \boldsymbol{r}_2, \tilde{\boldsymbol{c}}_i \rangle. \quad (33)$$

Let $\boldsymbol{a}_\ell$ denote the new atom selected by Oxx in the p-th iteration with $\boldsymbol{y}_1$ as input. By assumption, the atom selection is strict, *i.e.,*

$$|\langle \boldsymbol{r}_1, \tilde{\boldsymbol{c}}_\ell \rangle| > \max_{i \neq \ell} |\langle \boldsymbol{r}_1, \tilde{\boldsymbol{c}}_i \rangle|. \quad (34)$$

Taking the limit of (33) when $\varepsilon \to 0$, we obtain that for any i, $|\langle \boldsymbol{r}_{\mathcal{Q}}, \tilde{\boldsymbol{c}}_i \rangle|$ tends toward $|\langle \boldsymbol{r}_1, \tilde{\boldsymbol{c}}_i \rangle|$. (34) implies that when $\varepsilon < \varepsilon_{p-1}$ is sufficiently small,

$$|\langle \boldsymbol{r}_{\mathcal{Q}}, \tilde{\boldsymbol{c}}_\ell \rangle| > \max_{i \neq \ell} |\langle \boldsymbol{r}_{\mathcal{Q}}, \tilde{\boldsymbol{c}}_i \rangle|$$

by continuity of $|\langle \boldsymbol{r}_{\mathcal{Q}}, \tilde{\boldsymbol{c}}_i \rangle|$ ($i \neq \ell$) and $|\langle \boldsymbol{r}_{\mathcal{Q}}, \tilde{\boldsymbol{c}}_\ell \rangle|$ with respect to ε. Thus, Oxx with $\boldsymbol{y}(\varepsilon)$ as input selects $\boldsymbol{a}_\ell$ in the p-th iteration. ■

APPENDIX E
BAD RECOVERY CONDITION FOR BASIS PURSUIT

Contrary to the OMP analysis, the bad recovery analysis of basis pursuit is closely connected to the exact recovery analysis: in § III-E2, we argued that both analyses depend on the sign of the nonzero amplitudes, but not on the amplitude values [16, 23]. Here, we provide a more formal characterization of bad recovery for basis pursuit which is based on the Null Space Property (NSP) given in [32, Lemma 1]. The NSP is a sufficient and worst case necessary condition of exact recovery dedicated to all vectors whose support is equal to $\mathcal{Q}^\star$:

$$\forall \boldsymbol{x} \in \mathcal{N}(\boldsymbol{A}) \backslash \{\boldsymbol{0}\}, \sum_{i \in \mathcal{Q}^\star} |x_i| < \sum_{i \notin \mathcal{Q}^\star} |x_i| \quad \mathrm{NSP}(\boldsymbol{A}, \mathcal{Q}^\star)$$

where $\mathcal{N}(\boldsymbol{A}) = \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\}$ is the null space of $\boldsymbol{A}$.

Adapting the analysis of [32, Lemma 1], we introduce the following bad recovery condition.

Proposition 2

$$\forall \boldsymbol{\varepsilon} \in \{-1,1\}^k,\ \exists \boldsymbol{x} \in \mathcal{N}(\boldsymbol{A}),\ \sum_{i\in\mathcal{Q}^\star} \varepsilon_i x_i > \sum_{i\notin\mathcal{Q}^\star} |x_i| \qquad \text{BRC-BP}(\boldsymbol{A},\mathcal{Q}^\star)$$

is a necessary and sufficient condition of bad recovery by basis pursuit for any $\boldsymbol{x}^\star$ supported by $\mathcal{Q}^\star$.

This bad recovery condition reads as the intersection of as many conditions as possibilities for the sign vector $\boldsymbol{\varepsilon} \in \{-1,1\}^k$. We will see in the proof below that $\boldsymbol{\varepsilon}$ plays the role of the sign of the nonzero amplitudes, denoted by $\mathrm{sgn}(\boldsymbol{x}^\star) \in \{-1,1\}^k$. Therefore, the bad recovery condition is defined independently on each orthant related to some sign pattern $\boldsymbol{\varepsilon} \in \{-1,1\}^k$.

Proof: We first prove that BRC-BP is a sufficient condition for bad recovery for any $\boldsymbol{x}^\star$ supported by $\mathcal{Q}^\star$. For such a vector $\boldsymbol{x}^\star$, let $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}^\star$. Apply the BRC-BP condition for $\boldsymbol{\varepsilon}^\star \triangleq \mathrm{sgn}(\boldsymbol{x}^\star)$: there exists $\boldsymbol{x} \in \mathcal{N}(\boldsymbol{A})$ such that $\sum_{i\in\mathcal{Q}^\star} \varepsilon_i^\star x_i > \sum_{i\notin\mathcal{Q}^\star} |x_i|$. Because this inequality still holds when $\boldsymbol{x}$ is replaced by $\alpha\boldsymbol{x}$ (with $\alpha \neq 0$), we can freely re-scale $\boldsymbol{x}$ (*i.e.*, choose α small enough) so that for all $i \in \mathcal{Q}^\star$, $\mathrm{sgn}(x_i^\star - x_i) = \mathrm{sgn}(x_i^\star)$. Then, we have $|x_i^\star| = \varepsilon_i^\star x_i^\star = \varepsilon_i^\star(x_i^\star - x_i) + \varepsilon_i^\star x_i = |x_i^\star - x_i| + \varepsilon_i^\star x_i$ and

$$\begin{aligned} \|\boldsymbol{x}^\star\|_1 &= \sum_{i\in\mathcal{Q}^\star} |x_i^\star - x_i| + \sum_{i\in\mathcal{Q}^\star} \varepsilon_i^\star x_i \\ &> \sum_{i\in\mathcal{Q}^\star} |x_i^\star - x_i| + \sum_{i\notin\mathcal{Q}^\star} |x_i| = \|\boldsymbol{x}^\star - \boldsymbol{x}\|_1. \end{aligned}$$

Thus, $\boldsymbol{x}^\star$ cannot be a minimum ℓ_1 norm solution to $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$.

Now, let us prove that BRC-BP is also a necessary condition for bad recovery. Assume that $\boldsymbol{x}^\star$ is supported by $\mathcal{Q}^\star$ and basis pursuit with input $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}^\star$ yields output $\boldsymbol{x}^\star$. Because basis pursuit yields a minimum ℓ_1 norm solution to $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, we have for all $\boldsymbol{x} \in \mathcal{N}(\boldsymbol{A})$, $\|\boldsymbol{x}^\star - \boldsymbol{x}\|_1 \geqslant \|\boldsymbol{x}^\star\|_1$, *i.e.*,

$$\forall \boldsymbol{x} \in \mathcal{N}(\boldsymbol{A}),\ \sum_{i\notin\mathcal{Q}^\star} |x_i| \geqslant \sum_{i\in\mathcal{Q}^\star} |x_i^\star| - \sum_{i\in\mathcal{Q}^\star} |x_i^\star - x_i|. \tag{35}$$

Let $\boldsymbol{\varepsilon}^\star = \mathrm{sgn}(\boldsymbol{x}^\star)$ and $\rho = \min_{i\in\mathcal{Q}^\star} |x_i^\star|$. When $\|\boldsymbol{x}\|_\infty < \rho$, $x_i^\star - x_i$ and $x_i^\star$ are both of sign $\varepsilon_i^\star$ when $i \in \mathcal{Q}^\star$. Then, (35) yields:

$$\forall \boldsymbol{x} \in \mathcal{N}(\boldsymbol{A}),\ \|\boldsymbol{x}\|_\infty < \rho \Rightarrow \sum_{i\notin\mathcal{Q}^\star} |x_i| \geqslant \sum_{i\in\mathcal{Q}^\star} \varepsilon_i^\star x_i.$$

This condition also holds when $\|\boldsymbol{x}\|_\infty \geqslant \rho$ because it applies to $\rho\boldsymbol{x}/(2\|\boldsymbol{x}\|_\infty)$ whose ℓ_∞ norm is lower than ρ. We have shown the contrapositive of BRC-BP($\boldsymbol{A},\mathcal{Q}^\star$), *i.e.*, that BRC-BP($\boldsymbol{A},\mathcal{Q}^\star$) does not hold. ∎

We performed empirical tests for specific dictionaries of dimension ($m = 3, n = 5$) where $\mathcal{N}(\boldsymbol{A})$ is of dimension 2 and can be fully characterized. We checked that the BRC-BP property may indeed be fulfilled for $\mathrm{Card}\,[\mathcal{Q}^\star] = 2$.

References

[1] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation", *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[2] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries", *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[3] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition", in *Proc. 27th Asilomar Conf. on Signals, Systems and Computers*, Nov. 1993, vol. 1, pp. 40–44.

[4] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification", *Int. J. Control*, vol. 50, no. 5, pp. 1873–1896, Nov. 1989.

[5] B. K. Natarajan, "Sparse approximate solutions to linear systems", *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, Apr. 1995.

[6] T. Blumensath and M. E. Davies, "On the difference between Orthogonal Matching Pursuit and Orthogonal Least Squares", Tech. Rep., University of Edinburgh, Mar. 2007.

[7] A. J. Miller, *Subset Selection in Regression*, Chapman and Hall, London, UK, 2nd edition, Apr. 2002.

[8] S. F. Cotter, J. Adler, B. D. Rao, and K. Kreutz-Delgado, "Forward sequential algorithms for best basis selection", *IEE Proc. Vision, Image and Signal Processing*, vol. 146, no. 5, pp. 235–244, Oct. 1999.

[9] L. Rebollo-Neira and D. Lowe, "Optimized orthogonal matching pursuit approach", *IEEE Signal Process. Lett.*, vol. 9, no. 4, pp. 137–140, Apr. 2002.

[10] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems", *Proc. IEEE, invited paper (Special Issue "Applications of sparse representation and compressive sensing")*, vol. 98, no. 5, pp. 948–958, June 2010.

[11] M. E. Davies and Y. C. Eldar, "Rank awareness in joint sparse recovery", *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1135–1146, Feb. 2012.

[12] D. L. Donoho and Y. Tsaig, "Fast solution of ℓ_1-norm minimization problems when the solution may be sparse", *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4789–4812, Nov. 2008.

[13] M. A. Davenport and M. B. Wakin, "Analysis of orthogonal matching pursuit using the restricted isometry property", *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4395–4401, Sept. 2010.

[14] E. Liu and V. N. Temlyakov, "The orthogonal super greedy algorithm and applications in compressed sensing", *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2040–2047, Apr. 2012.

[15] Q. Mo and Y. Shen, "A remark on the restricted isometry property in orthogonal matching pursuit", *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3654–3656, June 2012.

[16] M. D. Plumbley, "On polar polytopes and the recovery of sparse representations", *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3188–3195, Sept. 2007.

[17] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit", *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[18] A. K. Fletcher and S. Rangan, "Orthogonal matching pursuit from noisy measurements: A new analysis", in *Neural Information Processing Systems*, Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, and A. Culotta, Eds. 2009, vol. 22, pp. 540–548, MIT Press.

[19] D. Donoho and M. Elad, "Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ^1 minimization", *Proc. Natl. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.

[20] T. Zhang, "Sparse recovery with orthogonal matching pursuit under RIP", *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 6215–6221, Sept. 2011.

[21] S. Foucart, "Stability and robustness of weak orthogonal matching pursuits", in *AMS Spring, Southeastern Conference*. 2011, Springer Proceedings in Mathematics.

[22] D. L. Donoho, "Neighborly polytopes and sparse solutions of underdetermined linear equations", Research Report, Stanford University, Stanford, CA, Dec. 2004.

[23] J.-J. Fuchs, "On sparse representations in arbitrary redundant bases", *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1341–1344, June 2004.

[24] D. A. Lorenz and D. Trede, "Greedy deconvolution of point-like objects", in *Signal Processing with Adaptive Sparse Structured Representations (SPARS workshop)*, Saint-Malo, France, Apr. 2009, pp. 1–6.

[25] C. Soussen, J. Idier, D. Brie, and J. Duan, "From Bernoulli-Gaussian deconvolution to sparse signal restoration", *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4572–4584, Oct. 2011.

[26] S. Bourguignon, C. Soussen, H. Carfantan, and J. Idier, "Sparse deconvolution: Comparison of statistical and deterministic approaches", in *IEEE Workshop Stat. Sig. Proc.*, Nice, France, June 2011.

[27] C. Dossal, *Estimation de fonctions géométriques et déconvolution*, Phd thesis, École Polytechnique, Palaiseau, France, Dec. 2005.

[28] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing", *Proceedings of the National Academy of Sciences of the USA*, pp. 1–6, Sept. 2009.

[29] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit", *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1094–1121, Feb. 2012.

[30] C. Dossal and S. Mallat, "Sparse spike deconvolution with minimum scale", in *Signal Processing with Adaptive Sparse Structured Representations (SPARS workshop)*, Rennes, France, Apr. 2005, pp. 1–4.

[31] R. Gribonval and M. Nielsen, "Beyond sparsity: Recovering structured representations by ℓ^1 minimization and greedy algorithms", *Adv. Comput. Math.*, vol. 28, no. 1, pp. 23–41, 2008.

[32] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases", *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.